

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:12:23 ; Search time 4552.47 Seconds  
(without alignments)  
17398.674 Million cell updates/sec

Title: US-09-899-718A-1  
Perfect score: 3785  
Sequence: 1 gttgggttcgctgttttc.....tggaccggtgttcgtcgac 3785

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.in.\*
- 32: em.htg.other.\*
- 33: em.htgo.in.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|-------|-------|--------|----|-------------|
| -----      |       |       |        |    |             |

|    |      |       |        |   |          |                    |
|----|------|-------|--------|---|----------|--------------------|
| 1  | 3785 | 100.0 | 3785   | 6 | AX349063 | Sequence           |
| 2  | 658  | 17.4  | 2805   | 8 | AB019622 | Triticum           |
| 3  | 544  | 14.4  | 2781   | 8 | AB029061 | Triticum           |
| 4  | 465  | 12.3  | 2781   | 8 | AB029063 | Triticum           |
| 5  | 312  | 8.2   | 2186   | 8 | TAWAXYSS | Wheat waxy         |
| 6  | 156  | 4.1   | 1801   | 8 | AF113843 | Triticum           |
| 7  | 93   | 2.5   | 2804   | 8 | AB029064 | Triticum           |
| 8  | 77   | 2.0   | 2028   | 8 | AF113844 | Triticum           |
| 9  | 77   | 2.0   | 2127   | 8 | AF163319 | Triticum           |
| 10 | 77   | 2.0   | 2793   | 8 | AB029062 | Triticum           |
| 11 | 77   | 2.0   | 2818   | 8 | AB019623 | Triticum           |
| 12 | 77   | 2.0   | 2834   | 8 | AF110373 | Triticum           |
| 13 | 72   | 1.9   | 2289   | 8 | AF286320 | Triticum           |
| 14 | 69   | 1.8   | 2893   | 8 | AF110375 | Aegilops           |
| 15 | 68   | 1.8   | 72     | 6 | AX349066 | Sequence           |
| 16 | 63   | 1.7   | 1605   | 8 | AF250137 | Triticum           |
| 17 | 63   | 1.7   | 2886   | 8 | AB019624 | Triticum           |
| 18 | 60   | 1.6   | 121    | 6 | AX325803 | Sequence           |
| 19 | 60   | 1.6   | 121    | 6 | AX325804 | Sequence           |
| 20 | 59   | 1.6   | 2826   | 8 | AF110374 | Aegilops           |
| 21 | 52   | 1.4   | 2311   | 8 | HWAXYR   | X07932 Barley mRNA |
| 22 | 52   | 1.4   | 5153   | 8 | HWAXYG   | X07931 Barley DNA  |
| 23 | 52   | 1.4   | 124050 | 8 | AF474373 | AF474373 Hordeum v |
| 24 | 47   | 1.2   | 81019  | 8 | AB073159 | AB073159 Arabidops |
| 25 | 44   | 1.2   | 121    | 6 | AX325799 | Sequence           |
| 26 | 44   | 1.2   | 121    | 6 | AX325800 | Sequence           |
| 27 | 41   | 1.1   | 121    | 6 | AX325823 | Sequence           |
| 28 | 41   | 1.1   | 121    | 6 | AX325824 | Sequence           |
| 29 | 39   | 1.0   | 2287   | 8 | SHU23945 | U33945 Sorghum bic |
| 30 | 39   | 1.0   | 4800   | 6 | AR106490 | AR106490 Sequence  |
| 31 | 39   | 1.0   | 4800   | 8 | MZEWAXY  | M24258 Maize amylo |
| 32 | 39   | 1.0   | 4800   | 8 | MZEWAXY  | X03935 Zea mays wa |
| 33 | 39   | 1.0   | 170949 | 2 | AC106104 | AC106104 Rattus no |
| 34 | 38   | 1.0   | 169048 | 2 | AC106675 | AC106675 Rattus no |
| 35 | 38   | 1.0   | 215352 | 2 | AC007305 | AC007305 Mus muscu |
| 36 | 37   | 1.0   | 121    | 6 | AX325835 | Sequence           |
| 37 | 37   | 1.0   | 121    | 6 | AX325836 | Sequence           |
| 38 | 37   | 1.0   | 121    | 6 | AX325839 | Sequence           |
| 39 | 37   | 1.0   | 121    | 6 | AX325840 | Sequence           |
| 40 | 37   | 1.0   | 121    | 6 | AX325843 | Sequence           |
| 41 | 37   | 1.0   | 121    | 6 | AX325844 | Sequence           |
| 42 | 37   | 1.0   | 18772  | 2 | AC109854 | AC109854 Rattus no |
| 43 | 37   | 1.0   | 69729  | 2 | AC102413 | AC102413 Mus muscu |
| 44 | 37   | 1.0   | 75756  | 9 | AC068601 | AC068601 Homo sapi |
| 45 | 37   | 1.0   | 134472 | 9 | AC005186 | AC005186 Human Chr |

## ALIGNMENTS

| RESULT   | 1          | AX349063   | Sequence 1 from Patent WO0202785.                         | 3785 bp | DNA | linear | PAT 06-FEB-2002 |
|----------|------------|------------|-----------------------------------------------------------|---------|-----|--------|-----------------|
| AX349063 | LOCUS      | AX349063   | Sequence 1 from Patent WO0202785.                         |         |     |        |                 |
|          | DEFINITION | AX349063   | AX349063                                                  |         |     |        |                 |
|          | ACCESSION  | AX349063   | AX349063                                                  |         |     |        |                 |
|          | VERSION    | AX349063.1 | GI:18615098                                               |         |     |        |                 |
|          | KEYWORDS   |            | bread wheat.                                              |         |     |        |                 |
|          | SOURCE     |            | Triticum aestivum                                         |         |     |        |                 |
|          | ORGANISM   |            |                                                           |         |     |        |                 |
|          | REFERENCE  |            | 1 (sites)                                                 |         |     |        |                 |
|          | AUTHORS    |            | Sprunck,S., Kluth,A., Becker,D., Luetticke,S. and Loez,H. |         |     |        |                 |
|          | TITLE      |            | Promoters of gene expression in plant caryopses           |         |     |        |                 |
|          | JOURNAL    |            | Patent: WO 0202785-A 1 10-JAN-2002;                       |         |     |        |                 |
|          | FEATURES   |            | Aventis CropScience GmbH (DE)                             |         |     |        |                 |
|          | Source     |            | Location/Qualifiers                                       |         |     |        |                 |
|          |            |            | 1. 3785                                                   |         |     |        |                 |
|          |            |            | /organism="Triticum aestivum"                             |         |     |        |                 |
|          |            |            | /db_xref="taxon:4565"                                     |         |     |        |                 |
|          | BASE COUNT |            | 945 a 980 c 899 g                                         |         |     |        |                 |
|          | ORIGIN     |            | 961 t                                                     |         |     |        |                 |

|                       |              |                                                                     |            |              |
|-----------------------|--------------|---------------------------------------------------------------------|------------|--------------|
| Query Match           | 100.0%       | Score 3785;                                                         | DB 6;      | Length 3785; |
| Best Local Similarity | 100.0%;      | Pred. No. 0;                                                        |            |              |
| Matches 3785;         | Conservative | 0;                                                                  | Mismatches | 0; Indels    |
|                       |              |                                                                     |            |              |
| Qy                    | 1            | gtttggtttcgcgttttttcatttcttcttcttaagggtgaatacaaatgacagtaat          | 60         |              |
|                       |              |                                                                     |            |              |
| Db                    | 1            | gtttgggtttcgctgttttttcattttccctttcttcttaagggtgaatacaaatgacagtaat    | 60         |              |
|                       |              |                                                                     |            |              |
| Qy                    | 61           | tcattatgtgaacagtcgagattcttgcgcaattatgtacaattcttttgaattgtt           | 120        |              |
|                       |              |                                                                     |            |              |
| Db                    | 61           | tcattatgtgttaacagtcgagattcttgcgcaattatgtacaattcttttgaattgtt         | 120        |              |
|                       |              |                                                                     |            |              |
| Qy                    | 121          | tgttcoatgtttatttcoatttcttcttaacttttttagggtaaaacaaatgccccaattca      | 180        |              |
|                       |              |                                                                     |            |              |
| Db                    | 121          | tgttttcattgtttttattttcattttcttcttaacttttttagggtaaaacaaatgccccaattca | 180        |              |
|                       |              |                                                                     |            |              |
| Qy                    | 181          | tctacacctaaagagaaatcagtttttactactagtttttcaagttttatttattgtattaaagt   | 240        |              |
|                       |              |                                                                     |            |              |
| Db                    | 181          | ttctpaccttaagagaaatcagtttttactactagtttttcaagttttatttattgtattaaagt   | 240        |              |
|                       |              |                                                                     |            |              |
| Qy                    | 241          | gttttttagtgggtttctcatattatgtatgcatgaattatgagggtgtgtgtoagtgt         | 300        |              |
|                       |              |                                                                     |            |              |
| Db                    | 241          | gttttttagtgggtttctcatattatgtatgcatgaattatgagggtgtgtgtoagtgt         | 300        |              |
|                       |              |                                                                     |            |              |
| Qy                    | 301          | gttaatatcacataagattattacacccatttttcagtcataaaattatgcaatttca          | 360        |              |
|                       |              |                                                                     |            |              |
| Db                    | 301          | gttaatatcacataagattattacacccatttttcagtcataaaattatgcaatttca          | 360        |              |
|                       |              |                                                                     |            |              |
| Qy                    | 361          | gtcaaaattgtgcgaaaactcttctcatttttatttttttttttttttttttttttttttaa      | 420        |              |
|                       |              |                                                                     |            |              |
| Db                    | 361          | gtacaaaattgtgcgaaaactcttctcattttttttttttttttttttttttttttttaa        | 420        |              |
|                       |              |                                                                     |            |              |
| Qy                    | 421          | gggtaaataccaaatgatactaaatttatccctcattttggaaatttcgttttgaaaattatgc    | 480        |              |
|                       |              |                                                                     |            |              |
| Db                    | 421          | gggttaataccaaatgatactaaatttatccctcattttggaaatttcgttttgaaaattatgc    | 480        |              |
|                       |              |                                                                     |            |              |
| Qy                    | 481          | tagtacacacttattcttgatattatgaaaagcgcaatttctgtgaagtttgttcatt          | 540        |              |
|                       |              |                                                                     |            |              |
| Db                    | 481          | tagtacacacttattcttgatattatgaaaagcgcaatttctgtgaagtttgttcatt          | 540        |              |
|                       |              |                                                                     |            |              |
| Qy                    | 541          | tcctgtatttttttcoattttcttcttcttggaagggtgaacactaatgcacaaattcat        | 600        |              |
|                       |              |                                                                     |            |              |
| Db                    | 541          | tcctgtatttttttcoattttttcttcttcttggaagggtgaacactaatgcacaaattcat      | 600        |              |
|                       |              |                                                                     |            |              |
| Qy                    | 601          | ctttgcttagaaaacttttagtattttgattgtgttttagttttttatttcattttgtttct      | 660        |              |
|                       |              |                                                                     |            |              |
| Db                    | 601          | ctttgctttagaaaacttttagtattttgattgtgttttagttttttatttcattttgtttct     | 660        |              |
|                       |              |                                                                     |            |              |
| Qy                    | 661          | tccttaagggaataaccaatgccactaatccattccattttagaanaatctctttattotla      | 720        |              |
|                       |              |                                                                     |            |              |
| Db                    | 661          | tccttttaagggaataaccaatgccactaatccattccattttagaanaatctctttattotla    | 720        |              |
|                       |              |                                                                     |            |              |
| Qy                    | 721          | caaaaactcaactttttatgcttattcgtgcataattataaaaagcacagtttcttatcta       | 780        |              |
|                       |              |                                                                     |            |              |
| Db                    | 721          | caaaaactcaactttttatgcttattcgtgcataattataaaaagcacagtttcttatcta       | 780        |              |
|                       |              |                                                                     |            |              |
| Qy                    | 781          | aatgtcgtgcaaaactttatcattttgtctaaattatttttcttagaataatgatgatacc       | 840        |              |
|                       |              |                                                                     |            |              |
| Db                    | 781          | aatgtcgtgcaaaactttatcattttgtctaaattatttttcttagaataatgatgatacc       | 840        |              |
|                       |              |                                                                     |            |              |
| Qy                    | 841          | aatgccactaatcattccgtgagacgcaatatgagggaatgcctacgtattattagtggt        | 900        |              |
|                       |              |                                                                     |            |              |
| Db                    | 841          | aatgccactaatcattccgtgagacgcaatatgagggaatgcctacgtattattagtggt        | 900        |              |
|                       |              |                                                                     |            |              |
| Qy                    | 901          | gtcgcatttttctctcaecgcatgggcataccctcacacatgcacacacaagcat             | 960        |              |
|                       |              |                                                                     |            |              |
| Db                    | 901          | gtcgcatttttctctcaecgcatgggcataccctcacacatgcacacacaagcat             | 960        |              |
|                       |              |                                                                     |            |              |
| Qy                    | 961          | acacacacatgagcactcacgcgagcacatgcatacaactgtgcgcacacacagcac           | 1020       |              |
|                       |              |                                                                     |            |              |
| Db                    | 961          | acacacacatgagcactcacgcgagcacatgcatacaactgtgcgcacacacagcac           | 1020       |              |
|                       |              |                                                                     |            |              |





```
gene
join(13..333,416..496,581..679,789..942,1068..1168,
1268..1621,1713..1892,1988..2179,2270..2356,2455..2583,
2677..2793)
/gene="waxy"
CDS
join(13..333,416..496,581..679,789..942,1068..1168,
1268..1621,1713..1892,1988..2179,2270..2356,2455..2583,
2677..2793)
/gene="waxy"
/codon_start=1
/product="starch synthase (GBSSI)"
/protein_id="BA77350.1"
/db_xref="GI:4760580"
/translation="MAALVTSQATSGTVLSVTDTRFRPFGQLRPNDAALGMRT
VGAAPKQSRKPHRFRCLSMVVRATGSGMNLVFGAEMAPWSKTGGLDVLGGL
PAAAMANGHRVMVISPRIYDOYKDAWDTSVISEIKVDRIYERVFHCYKRGVDRVFD
HPCFLVKRGKTEKIYGPAGTDYEDNQRFSLCQAALVPRILDLNNPHFSGPY
GEDVYFVNCNWHHTGLLACYLKSNGIYRTAKVAFCTHNTSYQGRSFDFAQLNL
PDRKSSFDIDGDKPVEGRKINMKKAGLLQADKVLTVSPYAEELISGEARGCLD
NIMRLTGITGIVNGMDVSEWDPIDKFLTVNYDVTALGKALKEALQAEGLPVDR
KVLPLVAFIGLEQKQPDVMIAPTEIVKEEDVQIVLLGTGKKKFERLKSVEEKPTT
KYVAVRFNAPLAHQMGADVLAVTSRPEPCGLIOLQGMRYGTPCACASTGGLVDTI
VEGKTGFHMRGLSDVCNVVPPADVKKVTTTLKRAVKVGTTPAYHEMVKMCIQDLSWK
GPAKNWEDVLLELGVGSEPGIIVGEIAPLALENVAAP"
BASE COUNT 577 a 817 c 826 g 585 t
ORIGIN
Query Match 17.4%; Score 658; DB 8; Length 2805;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 658; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3128 cctgcgcgcgcattcggcgtcgtcagctccacgtccacgtccgcacgtcctc 3187
Db 1 CCTGCGCGCCATGCGGCGCTGTCAGCTCCAGCTCCGACCTCCGGCACGCTC 60
QY 3188 agcgtcaccgacagattccggcgtccaggctttcaggccctgagcccccgaacccggc 3247
Db 61 AGCGTCACCGACAGATTCCGGCGCTCCAGGTTTTCAGGGCCCTGAGGCCCGGAACCCGGCG 120
QY 3248 gatgcgcgcctcgcattgagcactgctggagcagcagcccccacaaagcagagaaa 3307
Db 121 GATGCGGCGCTCGGATGAGGACTGTGGAGCGAGCGCCGCCCCCAAGACAGAGAAA 180
QY 3308 ccgcacccattccaccgcgctgcctctccatgctgctgcgcgcacggcagcgcgcgc 3367
Db 181 CCGCACCGATTCCGACCGCGCTGCTCTCCATGTTGTTGTCGCGCACGGCGCGGC 240
QY 3368 atgaacctcgttctgcgcgcgcagatggccctcctggagcagaactggcgcctcgcgc 3427
Db 241 ATGAACCTCGTGTTCGTGCGCGCGGAGATGGCGCCCTGGAGCAAGACTGGCGGCTCGGC 300
QY 3428 gacgcctcgggggctccctcccgcccatggcctgaagcttgccactgcctcttata 3487
Db 301 GACGTCCTCGGGGCGCTCCCGCGCCATGCCCTAGACTTGCCTCCTCTTATA 360
QY 3488 aatgtttctcctcagcactcctgcctgtacacaggggtgcggtgcctgcaggccaa 3547
Db 361 AATGTTTCTTCTCGACGCATGCTGCGCTTACAAAGGGTGCGGTGCTGTCGACAGGCAA 420
QY 3548 cggtaacgggtgatggtatctcccgcgctacgacagtaacagcagcctgggacac 3607
Db 421 CGGTACCGGTCATGTCATCTCCCGCGCTACGACAGTACAAAGACGCTGGGACAC 480
QY 3608 cagcgtatcctcaggtatatacccccacatgaattatcaaatccatcactcctcgc 3667
Db 481 CAGCGTCATCTCCAGGTATATATCCGCCACATGAATATCAATTCACATGCTCCTCG 540
QY 3668 acatttcgaagacttactgactgctgagctcgcagatcagatcaagctgtgacaggtta 3727
Db 541 ACATTTCGAAGACTTACTGACTGTGATCTCGCAGATCAAGTCTGTTGACAGGTA 600
QY 3728 cgagaggtgaggttacitccactctacaagcgggggtgagacggcgtgttgtcgac 3785
|||||
```

```
Db 601 CGAGAGGGTGAGTACTTCCACTGCTACAAAGCGGGGTGGACCGGTGTTGTCGAC 658
RESULT 3
AB029061
LOCUS 2781 bp DNA linear PLN 27-JAN-2001
DEFINITION Triticum turgidum subsp. dicoccoides waxy gene for starch synthase
(GBSSI), complete cds.
ACCESSION AB029061
VERSION AB029061.1 GI:6624280
KEYWORDS starch synthase (GBSSI).
SOURCE Triticum turgidum subsp. dicoccoides
ORGANISM Triticum turgidum subsp. dicoccoides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
REFERENCE 1 [sites]
AUTHORS Murai,J., Taira,T. and Ohta,D.
TITLE Isolation and characterization of the four waxy genes encoding the
granule-bound starch synthase in tetraploid wheats
JOURNAL Appl. Biol. Sci. (1999) In press
REFERENCE 2 [bases 1 to 2781]
AUTHORS Murai,J., Taira,T. and Ohta,D.
TITLE Direct Submission
JOURNAL Submitted (16-JUN-1999) Jun Murai, Osaka Prefecture University,
College of Agriculture; 1-1 Gakuen-cho, Sakai, Osaka 599-8531,
Japan (E-mail:junkiedmeter.plant.osakafu-u.ac.jp,
Tel:81-722-54-9409, Fax:81-722-54-9409)
FEATURES
Location/Qualifiers
1..2781
/organism="Triticum turgidum subsp. dicoccoides"
/sub_species="dicoccoides"
/db_xref="taxon:85692"
/gene="waxy"
join(1..321,404..484,569..667,777..930,1056..1156,
1256..1609,1701..1880,1976..2167,2258..2344,2443..2571,
2665..2781)
CDS
join(1..321,404..484,569..667,777..930,1056..1156,
1256..1609,1701..1880,1976..2167,2258..2344,2443..2571,
2665..2781)
/gene="waxy"
/codon_start=1
/product="starch synthase (GBSSI)"
/protein_id="BAA88509.1"
/db_xref="GI:6624281"
/translation="MAALVTSQATSGTVLSVTDTRFRPFGQLRPNDAALGMET
VGAAPKQSRKPHRFRCLSMVVRATGSGMNLVFGAEMAPWSKTGGLDVLGGL
PAAAMANGHRVMVISPRIYDOYKDAWDTSVISEIKVDRIYERVFHCYKRGVDRVFD
HPCFLVKRGKTEKIYGPAGTDYEDNQRFSLCQAALVPRILDLNNPHFSGPY
GEDVYFVNCNWHHTGLLACYLKSNGIYRTAKVAFCTHNTSYQGRSFDFAQLNL
PDRKSSFDIDGDKPVEGRKINMKAGILOADKVLTVSPYAEELISGEARGCLD
NIMRLTGITGIVNGMDVSEWDPIDKFLTVNYDVTALGKALKEALQAEGLPVDR
KYVAVRFNAPLAHQMGADVLAVTSRPEPCGLIOLQGMRYGTPCACASTGGLVDTI
VEGKTGFHMRGLSDVCNVVPPADVKKVTTTLKRAVKVGTTPAYHEMVKMCIQDLSWK
GPAKNWEDVLLELGVGSEPGIIVGEIAPLALENVAAP"
BASE COUNT 569 a 811 c 818 g 583 t
ORIGIN
Query Match 14.4%; Score 544; DB 8; Length 2781;
Best Local Similarity 99.7%; Pred. No. 1.6e-289;
Matches 644; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3140 atggcggtctcgtcagctccacgtccacgtccgcacgtcctcagcgtcacgcac 3199
Db 1 ATGGCGGCTCTGTCAGCTCCAGCTCCGACGCTCCGACCGCTCCTCAGGTCACCGAC 60
QY 3200 agattccgcgtccaggttttcaggcctgagccccgaacccgcggtcgcgcgc 3259
Db 61 AGATTCCGCGCTCAGGTTTTCAGGCTTGAGGCCCGCCGACCCCGGCGATGCGCGCTC 120
QY 3260 ggcatgagactgtcggagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3319
|||||
```



|                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                            |
|-----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------|
| QY                    | 3459                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | cgg 3461<br>                                                               |
| Db                    | 383                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | CCG 385                                                                    |
| RESULT                | 6                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                            |
| AF113843              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                            |
| LOCUS                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                            |
| DEFINITION            | AF113843 1801 bp mRNA linear PLN 20-APR-1999                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                            |
| ACCESSION             | Triticum aestivum granule-bound starch synthase precursor (Wx-Al)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                            |
| VERSION               | mRNA, Wx-Alb allele, complete cds.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                                            |
| KEYWORDS              | AF113843                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                            |
| SOURCE                | AF113843.1 GI:4586606                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                            |
| ORGANISM              | . bread wheat.<br>Triticum aestivum                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                            |
| REFERENCE             | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;<br>Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;<br>Poideae; Triticeae; Triticum.<br>1 (bases 1 to 1801)<br>Vrinten,P., Nakamura,T. and Yamamori,M.<br>Molecular characterization of waxy mutations in wheat<br>Mol. Gen. Genet. 261 (3), 463-471 (1999)<br>2 (bases 1 to 1801)<br>Vrinten,P.L. and Nakamura,T.<br>Direct Submission<br>Submitted (14-DEC-1998) Crop Breeding, Tohoku National Agricultural<br>Experiment Station, Shimo-Kuriyagawa, Morioka, Iwate 020-01, Japan                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                            |
| FEATURES              | Location/Qualifiers<br>1..1801<br>/organism="Triticum aestivum"<br>/cultivar="Waxy"<br>/db_xref="taxon:4565"<br>1..1801<br>/gene="Wx-Al"<br>/note="Non-functional Wx-Alb allele of waxy wheat;<br>spontaneously mutated Wx-Al gene"<br>32..1729<br>/gene="Wx-Alb"<br>/note="Waxy protein"<br>/codon_start=1<br>/product="granule-bound starch synthase precursor"<br>/protein_id="AAB26155.1"<br>/db_xref="GI:4588607"<br>/translation="MAALVTSLATSGTVLSVTDREPRPGFGLRPNPADAALGIRT<br>VGASAPKOSRKPHRRNRCLSWANGHRVMVISPRDQTKDAWDTSIVSEIKVDVRY<br>ERYVFHCYRGDFVFDHPCLEKRVGRDKTEKIYGPDACTDYDNQRFSLCQAQ<br>LEVPRIDLNNPFSGVDPVFCNDWHITGLLACYLKSNYSNGIYRTAKVAFCIT<br>HNTSYGRSFDDFAOLNIPDRKSSGFIDGYDKPVGEKRINMKAGILQADKVLTV<br>SPYAELLISGEARGCELNIIMRLTGITVNGMDVSEWDPIKDPLTVNDVTTALE<br>GNAKEARLOAEVGLPDVKPLVPTIGLEEOKGPDVMIAAPELVKEEDVOILVIG<br>TGKKFERLLKSVEEPKTRAVVRFNAPLAHMAGADVLAIVTSFEFGCLIQLOG<br>MRYTGCACASTGLVDITIVEGRTGHMRLSDCNVVEPADVKVVTTTLUKRAVKVWG<br>TPAYHEMVKNCMITQDLSSWKPAKNWEVDVLLLEGVGESEPGIVGEETAPLAENVAAP" |                                                                            |
| CDS                   | BASE COUNT 373 a 548 c 586 g 294 t                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                                            |
| ORIGIN                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                            |
| Query Match           | 4.1%; Score 156; DB 8; Length 1801;                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                            |
| Best Local Similarity | 100.0%; Pred. No. 3e-74;                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                            |
| Matches               | 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                            |
| Qy                    | 3109                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | tgtgtcaggtagccacacctgcgcgcccatggcggtctgttcacgtcccagctcgc 3168<br>     <br> |
| Db                    | 1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | TTGTTCAGGTAAGCACACCCTGGCGCGCATGGCGGCTCTGTGTCAGTCCCAGCTGC 60<br>            |
| Qy                    | 3169                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | cacctcggcacctcctcagctcacccagattccggcgctcaggttttcagggcc 3228<br>            |
| Db                    | 61                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | CACCTCGGCACTGCTCTCAGCGTCACGACAGATTCCGGCGTTCAGGCTTTTAGGGCCT 120<br>         |
| Qy                    | 3229                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | gagccccgggaaccggcgatgcggcgctcggcat 3264<br>                                |

```

Db 121 GAGGCCCGGGAACCGCGGATGCGCGCTGGCAT 156

RESULT 7
LOCUS AB029064
DEFINITION Triticum durum waxy gene for starch synthase (GBSSI), complete cds.
ACCESSION AB029064
VERSION AB029064.1 GI:6624286
KEYWORDS Starch synthase (GBSSI).
SOURCE Triticum durum DNA
ORGANISM Triticum turgidum subsp. durum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
1 (sites)
Murai,J., Taira,T. and Ohta,D.
Isolation and characterization of the four Waxy genes encoding the
granule-bound starch synthase in tetraploid wheats
Appl. Biol. Sci. (1999) In press
2 (bases 1 to 2804)
Murai,J., Taira,T. and Ohta,D.
Direct Submission
Submitted (16-JUN-1999) Jun Murai, Osaka Prefecture University,
College of Agriculture; 1-1 Gakuen-cho, Sakai, Osaka 599-8531,
Japan (E-mail:junki@emeter.plant.osakafu-u.ac.jp,
Tel:81-722-54-9409, Fax:81-722-54-9409)
Location/Qualifiers
1..2804
/organism="Triticum turgidum subsp. durum"
/db_xref="taxon:4567"
join(1..321,418..498,590..688,799..952,1093..1193,
1293..1646,1731..1910,1993..2184,2265..2351,2439..2567,
2688..2804)
/gene="waxy"
join(1..321,418..498,590..688,799..952,1093..1193,
1293..1646,1731..1910,1993..2184,2265..2351,2439..2567,
2688..2804)
/gene="waxy"
/codon_start=1
/product="starch synthase (GBSSI)"
/db_xref="GI:6624287"
/translation="MAALVTSQATSGVLGITDRFRAGFGVVRPNPADAALVMRT
IGASAPKSRKAHRGRSLSMVVRATSGGMNLVFGAEMAPWSKTKGGLDVLGGL
PPAAMANGHRVMVISPRIYDQYKDAWDTSVSEIKVADEYRVFHCYKRGVDRFVD
HPCELEKVRGKTEKIYGPDAQTDYEDNQRFSLCQAALAPRTLDLNNPNYFSGPY
GDEVVFCNDWHTGLACLYKSNQNGIYMTAKVAFCHINISYQGRSFDFDFAQLNL
PDRFKSSDFIDYDKPVEGRKINMKAGILQADKVLTVSPYAEELISGARCELD
NIMRLTGITGVNGMDVSEWDPDKFLAVNYDVTITALEGKALKEALQAEVGLPVR
KVLPAFTIGLEOKGPDVMTAAIPEILKEEDVQIVLLGTGKKKFERLLKSVEEKFPN
KYRVAVRNAPLAHOMMAGADVLAVTSRFEPCGLIQGMRYGTPCASCSTGGLVDTI
VEGKTGFHMRGLSDCNVPEADVKVVTTLKRAVKKVGTTPAYHGMVKNCMIQDLSWK
GPAKMWEDVLELGESEPGVGEIAPLAMENVAAP"

BASE COUNT 577 a 823 c 824 g 580 t
ORIGIN

Query Match 2.5%; Score 93; DB 8; Length 2804;
Best Local Similarity 100.0%; Pred. No. 2.7e-39;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3322 ccggcggtccctccatggtgtagcgcccaaggcgagcgagcctcgtgtt 3381
|||||
Db 183 CCGCGGTCCTCTCATGTTGGTGGCGCCACGCGGCGCATGACCTCGTGT 242

QY 3382 cgtcgccgagagtagcgccctggagcaagac 3414
|||||
Db 243 CCGCGCGCGAGATGCGCCCTGGAGCAAGAC 275

RESULT 8
AF113844
LOCUS AF163319
DEFINITION Triticum aestivum granule-bound starch synthase I (Wx-DI) mRNA,
Wx-DIa allele, complete cds.
ACCESSION AF163319
VERSION AF163319.1 GI:6969977

```



**AUTHORS** Mirai, J., Taira, T. and Ohta, D.  
**TITLE** Isolation and characterization of the three waxy genes encoding the granule-bound starch synthase in hexaploid wheat

**JOURNAL** Gene 234 (1), 71-79 (1999)  
**MEDLINE** 99321800

**REFERENCE** 2 (bases 1 to 2818)  
**AUTHORS** Mirai, J., Taira, T. and Ohta, D.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-NOV-1998) Jun Murai, Osaka Prefecture University, College of Agriculture; 1-1 Gakuen-cho, Sakai, Osaka 599-8531, Japan (E-mail: junkiedemeter.plant.osakafu-u.ac.jp, Tel: 81-722-54-9409, Fax: 81-722-54-9409)

**FEATURES**  
 source location/Qualifiers  
 1. .2818  
 /organism="Triticum aestivum"  
 /db\_xref="taxon:4565"  
 join(13..336,436..516,605..703,817..970,1104..1204,1274..1627,1720..1899,1986..2177,2262..2348,2446..2574,2690..2806)  
 /gene="waxy"  
 join(13..336,436..516,605..703,817..970,1104..1204,1274..1627,1720..1899,1986..2177,2262..2348,2446..2574,2690..2806)  
 /gene="waxy"  
 /codon\_start=1  
 /product="starch synthase (GBSSI)"  
 /protein\_id="BAA77351.1"  
 /db\_xref="GI:4760582"  
 /translation="MAALVTSQATSGVILGIDRRPRAGFGVGRPSADALGMRT TGASAPKQSRKAHRCRCLSMVYVATSGAGMLVFGAEMAPWSKTGGDLVIGG LPPAMAANGHRVMSIPRYDOYKDAWDTSVSEIKVADEYERVYFHCYKRGVDVFFV DHPCFLEKVRGKTKIYDPDAGTDYEDNQLRFSLLCOALAPRILDLNNPYSFGP YGDDVFVNCNDWHTGLLACYLKSNSQSGIYRTAKVAFCHNISYQGRSFDDFAQLN LPRDKFSDFIDGDKPVEGRKINMKAGILOADKVLTVSPYAEELISGEARGCEL DNIMRLGTIGVINGMDVSEWDPDKNLFVNYVTALSGKVLNKEALQAEVGLPVD RKVPLVAFIRLEQKQPDVMTAAIPEILKEEDYQVILVLTGKKRERLLKSVEEKFP SKRVAVRFPNAPLQHMAGADVLAVTSREPCGLIQLOQMRGYTPCACASTGGLVDT IMEGKTGFHMRJSLSDCNVVEPADVKVVTTLKRAVKVGTGPAYHEMVKMCMQIDLWS KGPANWEDVLELGVGSEPGVIEEIAPLAMENVAAP"

**BASE COUNT** 579 a 846 c 828 g 565 t  
**ORIGIN**

**Query Match** 2.0%; Score 77; DB 8; Length 2818;  
**Best Local Similarity** 100.0%; Pred. No. 2.1e-30;  
**Matches** 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3538 tgcaggccaacggtaccgggtcatgttcattccccgcgtacaccagtagcaaggagc 3597  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 431 TGCAGGCCAACGTCACCGGGTCATGTCATCTCCCGCGCTACGACGAGTACAAGGACG 490  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 3598 cctgggacacacagcgctc 3614  
 ||||||||||||||||||||  
 Db 491 CCTGGGACACGCGTC 507  
 ||||||||||||||||

**RESULT 12**  
**LOCUS** AF110373  
**DEFINITION** Triticum monococcum granule-bound starch synthase WX-Tma protein (wx-Tma) gene, complete cds.  
**ACCESSION** AF110373  
**VERSION** AF110373.1  
**KEYWORDS** GI:6318537  
**SOURCE** Triticum monococcum.  
**ORGANISM** Triticum monococcum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.  
 1 (bases 1 to 2834)  
 Yan, L., Bhawe, M., Fairclough, R., Konic, C., Rahman, S. and Appels, R.  
**REFERENCE** The genes encoding granule-bound starch synthases at the waxy loci of the A, B and D progenitors of common wheat

**JOURNAL** Genome (1999) In press  
**REFERENCE** 2 (bases 1 to 2834)  
**AUTHORS** Yan, L., Bhawe, M., Fairclough, R., Konic, C., Rahman, S. and Appels, R.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (02-DEC-1998) School of Life Sciences and Technology, Victoria University of Technology, Werribee Campus, PO Box 14428 MCMC, Melbourne, Victoria 8001, Australia

**FEATURES**  
 source location/Qualifiers  
 1. .2834  
 /organism="Triticum monococcum"  
 /cultivar="AUS 22986"  
 /db\_xref="taxon:4568"  
 join(32..355,444..524,613..711,830..983,1127..1227,1309..1662,1752..1931,2013..2204,2288..2374,2474..2602,2718..2834)  
 /gene="wx-Tma"  
 <32..>2834  
 /gene="wx-Tma"  
 /note="waxy"  
 join(32..355,444..524,613..711,830..983,1127..1227,1309..1662,1752..1931,2013..2204,2288..2374,2474..2602,2718..2834)  
 /gene="wx-Tma"  
 /note="GBSSI"  
 /codon\_start=1  
 /product="granule-bound starch synthase WX-Tma protein"  
 /protein\_id="AAF06936.1"  
 /db\_xref="GI:6318538"  
 /translation="MAALVTSQATSGVILGIDRRPRAGFGVGRPSADALGMRT VGASAPKQSRKAHRCRCLSMVYVATSGAGMLVFGAEMAPWSKTGGDLVIGG LPPAMAANGHRVMSIPRYDOYKDAWDTSVSEIKVADEYERVYFHCYKRGVDVFFV DHPCFLEKVRGKTKIYDPDAGTDYEDNQLRFSLLCOALAPRILDLNNPYSFGP YGDDVFVNCNDWHTGLLACYLKSNSQSGIYRTAKVAFCHNISYQGRSFDDFAQLN LPRDKFSDFIDGDKPVEGRKINMKAGILOADKVLTVSPYAEELISGEARGCEL DNIMRLGTIGVINGMDVSEWDPDKNLFVNYVTALSGKVLNKEALQAEVGLPVD RKVPLVAFIRLEQKQPDVMTAAIPEILKEEDYQVILVLTGKKRERLLKSVEEKFP SKRVAVRFPNAPLQHMAGADVLAVTSREPCGLIQLOQMRGYTPCACASTGGLVDT IMEGKTGFHMRJSLSDCNVVEPADVKVVTTLKRAVKVGTGPAYHEMVKMCMQIDLWS KGPANWEDVLELGVGSEPGVIEEIAPLAMENVAAP"

**BASE COUNT** 578 a 834 c 831 g 591 t  
**ORIGIN**

**Query Match** 2.0%; Score 77; DB 8; Length 2834;  
**Best Local Similarity** 100.0%; Pred. No. 2.1e-30;  
**Matches** 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3538 tgcaggccaacggtaccgggtcatgttcattccccgcgtacaccagtagcaaggagc 3597  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 439 TGCAGGCCAACGTCACCGGGTCATGTCATCTCCCGCGCTACGACGAGTACAAGGACG 498  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 3598 cctgggacacacagcgctc 3614  
 ||||||||||||||||||||  
 Db 499 CCTGGGACACGCGTC 515  
 ||||||||||||||||

**RESULT 13**  
**LOCUS** AF286320  
**DEFINITION** Triticum aestivum granule bound starch synthase I (gbss1) gene, complete cds.  
**ACCESSION** AF286320  
**VERSION** AF286320.1  
**KEYWORDS** GI:11037535  
**SOURCE** bread wheat.  
**ORGANISM** Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.  
 1 (bases 1 to 2289)  
 McCue, K.F., Harkman, W.J., Tanaka, C.K. and Anderson, O.D.  
**REFERENCE** Starch Branching Enzymes Sbe1 and Sbe2 from Wheat (Triticum aestivum cv. Cheyenne): Molecular Characterization, Developmental



|                                                                      |                                                                    |  |  |                                                                                |  |  |
|----------------------------------------------------------------------|--------------------------------------------------------------------|--|--|--------------------------------------------------------------------------------|--|--|
| JOURNAL<br>REFERENCE<br>AUTHORS<br>TITLE<br>JOURNAL                  | Expression, and Homolog Assignment by Differential PCR             |  |  | Unpublished                                                                    |  |  |
|                                                                      | 2 (bases 1 to 2289)                                                |  |  | McCue,K.F. and Anderson,O.D.                                                   |  |  |
|                                                                      | Direct Submission                                                  |  |  | Submitted (11-JUL-2000) United States Department of Agriculture,               |  |  |
|                                                                      | Agricultural Research Service, 800 Buchanan Street, Albany, CA     |  |  | 94710-1105, USA                                                                |  |  |
| FEATURES<br>source                                                   | Location/Qualifiers                                                |  |  | Location/Qualifiers                                                            |  |  |
|                                                                      | 1. .2289                                                           |  |  | 1. .2893                                                                       |  |  |
|                                                                      | /organism="Triticum aestivum"                                      |  |  | /organism="Aegilops tauschii"                                                  |  |  |
|                                                                      | /cultivar="Cheyenne"                                               |  |  | /cultivar="CPI 110799"                                                         |  |  |
| gene                                                                 | /db_xref="taxon:4565"                                              |  |  | /db_xref="taxon:37682"                                                         |  |  |
|                                                                      | /tissue_type="endosperm"                                           |  |  | join(<32..352,443..523,620..718,823..976,1128..1228,                           |  |  |
|                                                                      | 204..2021                                                          |  |  | 1370..1723,1809..1988,2071..2262,2347..2433,2532..2660,                        |  |  |
|                                                                      | 204..2021                                                          |  |  | 2777..>2893)                                                                   |  |  |
| CDS                                                                  | /gene="Gbssi"                                                      |  |  | /gene="wx-Ttd"                                                                 |  |  |
|                                                                      | 204..2021                                                          |  |  | <32..>2893                                                                     |  |  |
|                                                                      | /gene="Gbssi"                                                      |  |  | /gene="wx-Ttd"                                                                 |  |  |
|                                                                      | /EC_number="2.4.1.21"                                              |  |  | /note="waxy"                                                                   |  |  |
|                                                                      | /note="glucosyltransferase; ADP glucose glucosyl                   |  |  | join(32..352,443..523,620..718,823..976,1128..1228,                            |  |  |
|                                                                      | transferase; waxy; wx"                                             |  |  | 1370..1723,1809..1988,2071..2262,2347..2433,2532..2660,                        |  |  |
|                                                                      | /codon_start=1                                                     |  |  | 2777..2893)                                                                    |  |  |
|                                                                      | /product="granule bound starch synthase I"                         |  |  | /gene="GBSSI"                                                                  |  |  |
|                                                                      | /protein_id="AAG27624.1"                                           |  |  | /codon_start=1                                                                 |  |  |
|                                                                      | /db_xref="GI:11037536"                                             |  |  | /product="granule-bound starch synthase WX-Ttd protein"                        |  |  |
|                                                                      | /translation="MAALVTSQATSGTVLGIITDRFRAGFGQVRPSADALGMRT             |  |  | /protein_id="AAF06938.1"                                                       |  |  |
|                                                                      | TGSAAPKQGRKAHGRTRCLSMVVRATSGGMNLVFGAEMAPWSTKGGIDVLGG               |  |  | /translation="MAALVTSQATSGTVLGIITDRFRAGFGQVRPSADALGMRT                         |  |  |
|                                                                      | LPAMAANGHRVWISPRDQDKADMTSVSEIKVADYERVRVHCYKRGVDRVFF                |  |  | VGSAAPTQSRKAHGRTRCLSMVVRATSGGMNLVFGAEMAPWSTKGGIDVLGG                           |  |  |
|                                                                      | DHPFLEKVRGKTKIYIGPDAGTDYEDNQLRFSLCQAALAPRLLDNNPYSGP                |  |  | PPMAAANGHRVWISPRDQDKADMTSVSEIKVADYERVRVHCYKRGVDRVFF                            |  |  |
|                                                                      | YGDVFVNCNDHTGLLACYLKSNTQSSGIYRTKAVFCIHNSIOGRRFDDFAQLN              |  |  | HPCFLEKVRGKTKIYIGPDAGTDYEDNQLRFSLCQAALAPRLLDNNPYSGP                            |  |  |
|                                                                      | LDPFKSSFFIDYQKPVGRKINMKAGIILQADKVLTVSPYAEELISGBARCEL               |  |  | GEDVFCVNCNDHTGLLACYLKSNTQSSGIYRTKAVFCIHNSIOGRRFDDFAQLN                         |  |  |
|                                                                      | DNTMLRTGIGTVNGMDVSEDPKDKFLAANDVDTALEGKALKEALQAEVGLPD               |  |  | PDRKSSFFIDYQKPVGRKINMKAGIILQADKVLTVSPYAEELISGBARCEL                            |  |  |
|                                                                      | RYPLVAFIRGLEQKGPDMIAAIPKILKEEDVQIVLLTGKTKKPKLKSVEKFPV              |  |  | DNTMLRTGIGTVNGMDVSEDPKDKFLAANDVDTALEGKALKEALQAEVGLPD                           |  |  |
|                                                                      | SKRVAVRENAPLAHOMAGADVLAVTSRFEPCGLIQLOGMRYGTPCACASTGGGLVD           |  |  | KVPLVAFIRGLEQKGPDMIAAIPKILKEEDVQIVLLTGKTKKPKLKSVEKFPV                          |  |  |
|                                                                      | IMEGKTGFHMGHLSVDCNVVEPADVKVVTTLKRAVKVVGTPAYHEMVKNCMIQDLSW          |  |  | KRVAVRENAPLAHOMAGADVLAVTSRFEPCGLIQLOGMRYGTPCACASTGGGLVD                        |  |  |
| BASE COUNT                                                           | 480 a 684 c 735 g 390 t                                            |  |  | 596 a 838 c 841 g 618 t                                                        |  |  |
|                                                                      | ORIGIN                                                             |  |  | ORIGIN                                                                         |  |  |
|                                                                      | KGPAKNMEDVLELGESEPGVIGEEIAPLAMENVAAP"                              |  |  | VEGKTGFHMGHLSVDCNVVEPADVKVVTTLKRAVKVVGTPAYHEMVKNCMIQDLSWK                      |  |  |
|                                                                      |                                                                    |  |  | KGPAKNMEDVLELGESEPGVIGEEIAPLAMENVAAP"                                          |  |  |
| Query Match                                                          | 1.98; Score 72; DB 8; Length 2289;                                 |  |  | 1.8%; Score 69; DB 8; Length 2893;                                             |  |  |
|                                                                      | Best Local Similarity 100.0%; Pred. No. 1.2e-27;                   |  |  | Best Local Similarity 100.0%; Pred. No. 5.7e-26;                               |  |  |
|                                                                      | Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;        |  |  | Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;                    |  |  |
|                                                                      |                                                                    |  |  |                                                                                |  |  |
| QY 3543 gccaacgtaccgggtcatgtcatctccccgcgtaccagcagtcacagacgcctgg 3602 |                                                                    |  |  |                                                                                |  |  |
|                                                                      | Db 528 GCCAACGGTGACCGGTCATGTCATCTCCCGCGCTACGACGACGACGCTGG 587      |  |  | QY 3119 tagccacacccctgcgcgcgcctatgcgcgcctctggtcagctccacgtccacacccctgcgcgc 3178 |  |  |
|                                                                      |                                                                    |  |  |                                                                                |  |  |
|                                                                      |                                                                    |  |  | Db 11 TAGCCACACCCCTGCGCGCGCCATGCGCGCTCTGTCACGCTCCCGACCTCCCGGC 70               |  |  |
| QY 3603 gacaccagcgtc 3614                                            |                                                                    |  |  | QY 3179 accgtctctc 3187                                                        |  |  |
|                                                                      | Db 588 GACACGAGGTC 599                                             |  |  |                                                                                |  |  |
|                                                                      |                                                                    |  |  | Db 71 ACCGTCTCTC 79                                                            |  |  |
|                                                                      |                                                                    |  |  |                                                                                |  |  |
| RESULT 14                                                            | AF110375                                                           |  |  | AX349066                                                                       |  |  |
|                                                                      | LOCUS                                                              |  |  | LOCUS                                                                          |  |  |
|                                                                      | DEFINITION                                                         |  |  | DEFINITION                                                                     |  |  |
|                                                                      | Aegilops tauschii granule-bound starch synthase wx-ttd protein     |  |  | Sequence 4 from Patent WO0202785.                                              |  |  |
| ACCESSION                                                            | AF110375                                                           |  |  | AX349066                                                                       |  |  |
|                                                                      | VERSION                                                            |  |  | AX349066.1                                                                     |  |  |
|                                                                      | KEYWORDS                                                           |  |  | KEYWORDS                                                                       |  |  |
|                                                                      | SOURCE                                                             |  |  | SOURCE                                                                         |  |  |
| ORGANISM                                                             | Aegilops tauschii                                                  |  |  | Aegilops tauschii                                                              |  |  |
|                                                                      | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; |  |  | synthetic construct.                                                           |  |  |
|                                                                      | Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;         |  |  | artificial sequence.                                                           |  |  |
|                                                                      | Poideae; Triticeae; Aegilops.                                      |  |  | 1 (sites)                                                                      |  |  |
| REFERENCE                                                            | 1 (bases 1 to 2893)                                                |  |  | Sprunck,S., Kluth,A., Becker,D., Luetticke,S. and Loerz,H.                     |  |  |
|                                                                      | Yan,L., Bhawe,M., Fairclough,R., Konic,C., Rahman,S. and Appels,R. |  |  | Promoters of gene expression in plant carypsops                                |  |  |
|                                                                      | The genes encoding granule-bound starch synthases at the waxy loci |  |  | Patent: WO 0202785-A 4 10-JAN-2002;                                            |  |  |
|                                                                      | of the A, B and D progenitors of common wheat                      |  |  | Aventis CropScience GmbH (DE)                                                  |  |  |
| JOURNAL                                                              | Genome (1999) In press                                             |  |  | Location/Qualifiers                                                            |  |  |
|                                                                      | 2 (bases 1 to 2893)                                                |  |  | 1. .72                                                                         |  |  |
|                                                                      | Yan,L., Bhawe,M., Fairclough,R., Konic,C., Rahman,S. and Appels,R. |  |  | /organism="synthetic construct"                                                |  |  |
|                                                                      |                                                                    |  |  | /db_xref="taxon:32630"                                                         |  |  |
| AUTHORS                                                              |                                                                    |  |  | /note="Oligonucleotide"                                                        |  |  |
|                                                                      |                                                                    |  |  |                                                                                |  |  |
|                                                                      |                                                                    |  |  |                                                                                |  |  |
|                                                                      |                                                                    |  |  |                                                                                |  |  |

Search completed: July 31, 2002, 16:11:39  
Job time: 14356 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:26:33 ; Search time 79.86 Seconds  
(without alignments)  
11641.908 Million cell updates/sec

Title: US-09-899-718A-1

Perfect score: 3785

Sequence: 1 gtttggttcgcgtgttttc.....tgaccgcgtgttcgtcgac 3785

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size: 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description       |
|------------|-------|-------------|--------|-------|-------------------|
| 1          | 39    | 1.0         | 2267   | 4     | US-08-679-645-25  |
| 2          | 39    | 1.0         | 4800   | 3     | US-08-941-445A-4  |
| 3          | 31    | 0.8         | 43795  | 3     | US-08-742-185-101 |
| 4          | 25    | 0.7         | 252    | 2     | US-08-623-906A-1  |
| 5          | 25    | 0.7         | 2542   | 3     | US-08-941-445A-6  |
| 6          | 24    | 0.6         | 39     | 4     | US-08-860-038-23  |
| 7          | 24    | 0.6         | 39     | 4     | US-09-580-923-23  |
| 8          | 24    | 0.6         | 1865   | 4     | US-08-370-253-5   |
| 9          | 24    | 0.6         | 3211   | 2     | US-08-574-959A-8  |
| 10         | 24    | 0.6         | 3211   | 4     | US-09-357-014-8   |
| 11         | 24    | 0.6         | 3901   | 2     | US-08-574-959A-6  |
| 12         | 24    | 0.6         | 3901   | 4     | US-09-357-014-6   |
| 13         | 24    | 0.6         | 5394   | 3     | US-08-688-376-1   |
| 14         | 23    | 0.6         | 2621   | 2     | US-08-553-619B-8  |
| 15         | 23    | 0.6         | 3387   | 1     | US-08-261-822A-5  |
| 16         | 23    | 0.6         | 3387   | 5     | PCT-US95-07744A-5 |
| 17         | 22    | 0.6         | 25     | 3     | US-08-894-511-5   |
| 18         | 22    | 0.6         | 25     | 4     | US-08-860-038-1   |
| 19         | 22    | 0.6         | 25     | 4     | US-08-860-038-12  |
| 20         | 22    | 0.6         | 25     | 4     | US-09-580-923-1   |
| 21         | 22    | 0.6         | 36     | 3     | US-08-941-445A-29 |
| 22         | 22    | 0.6         | 51     | 4     | US-09-580-923-33  |
| 23         | 22    | 0.6         | 57     | 3     | US-08-894-511-3   |
| 24         | 22    | 0.6         | 57     | 3     | US-08-894-511-4   |
| 25         | 22    | 0.6         | 57     | 3     | US-08-894-511-18  |
| 26         | 22    | 0.6         | 58     | 4     | US-08-860-038-9   |
| 27         | 22    | 0.6         | 58     | 4     | US-08-860-038-10  |

|      |    |     |     |   |                  |
|------|----|-----|-----|---|------------------|
| 28   | 22 | 0.6 | 58  | 4 | US-09-580-923-9  |
| c 29 | 22 | 0.6 | 58  | 4 | US-09-580-923-10 |
| c 30 | 22 | 0.6 | 62  | 6 | 5457089-7        |
| 31   | 22 | 0.6 | 66  | 3 | US-08-894-511-16 |
| c 32 | 22 | 0.6 | 66  | 3 | US-08-894-511-17 |
| 33   | 22 | 0.6 | 66  | 4 | US-08-860-038-20 |
| c 34 | 22 | 0.6 | 66  | 4 | US-08-860-038-21 |
| 35   | 22 | 0.6 | 66  | 4 | US-09-580-923-20 |
| c 36 | 22 | 0.6 | 66  | 4 | US-09-580-923-21 |
| 37   | 22 | 0.6 | 188 | 1 | US-08-115-497-21 |
| c 38 | 22 | 0.6 | 188 | 1 | US-08-466-670-21 |
| c 39 | 22 | 0.6 | 188 | 2 | US-08-291-011-1  |
| c 40 | 22 | 0.6 | 188 | 4 | US-09-266-065-1  |
| c 41 | 22 | 0.6 | 405 | 2 | US-08-299-074A-1 |
| c 42 | 22 | 0.6 | 405 | 4 | US-09-399-773-1  |
| c 43 | 22 | 0.6 | 550 | 4 | US-09-306-042-2  |
| c 44 | 22 | 0.6 | 678 | 4 | US-09-227-357-81 |
| c 45 | 22 | 0.6 | 884 | 1 | US-08-365-981-4  |

## ALIGNMENTS

RESULT 1  
US-08-679-645-25  
; Sequence 25, Application US/08679645  
; Patent No. 6350934  
; GENERAL INFORMATION:  
; APPLICANT: Zwick, Michael G.  
; APPLICANT: Edington, Brent E.  
; APPLICANT: McSwiggen, James A.  
; APPLICANT: Merlo, Patricia Ann Owens  
; APPLICANT: Guo, Lining  
; APPLICANT: Skokut, Thomas A.  
; APPLICANT: Young, Scott A.  
; APPLICANT: Folkerts, Otto  
; APPLICANT: Merlo, Donald J.  
; TITLE OF INVENTION: COMPOSITION AND METHODS FOR  
; TITLE OF INVENTION: MODULATION OF GENE EXPRESSION  
; TITLE OF INVENTION: IN PLANTS  
; NUMBER OF SEQUENCES: 1263  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/679,645  
; FILING DATE: July 12, 1996  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/001,135  
; FILING DATE: July 13, 1995  
; APPLICATION NUMBER: 08/300,726  
; FILING DATE: September 2, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 219/247  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:  
LENGTH: 2267 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-679-645-25

Query Match 1.0%; Score 39; DB 4; Length 2267;  
Best Local Similarity 100.0%; Pred. No. 6.8e-09;  
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3576 cgctacgaccagctacagacgacctgggacacacgctc 3614  
|||||

DB 543 CGCTACGACCAGCTACAGACGACCTGGGACACACGCTC 581

## RESULT 2

US-08-941-445A-4  
Sequence 4, Application US/08941445A

Patent No. 6107060

GENERAL INFORMATION:

APPLICANT: Keeling, Peter

APPLICANT: Guan, Hanning

TITLE OF INVENTION: Starch Encapsulation

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Greenlee, Winner and Sullivan, P.C.

STREET: 5370 Manhattan Circle

CITY: Boulder

STATE: CO

COUNTRY: US

ZIP: 80303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/941,445A

FILING DATE: 30-SEP-1997

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/026,855

FILING DATE: 30-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Winner, Ellen P

REGISTRATION NUMBER: 28,547

REFERENCE/DOCKET NUMBER: 89-97

TELEPHONE: (303) 499-8080

TELEFAX: (303) 499-8089

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 4800 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: not relevant

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ORGANISM: Zea mays

FEATURE:

NAME/KEY: CDS

LOCATION: Join(1449..1553, 1685..1765, 1860..1958, 2055

LOCATION: ..2144, 2226..2289, 2413..2513, 2651..2760, 2858

LOCATION: ..3101, 3212..3394, 3490..3681, 3793..3879, 3977

LOCATION: ..4105, 4227..4343)

US-08-941-445A-4

Query Match 1.0%; Score 39; DB 3; Length 4800;  
Best Local Similarity 100.0%; Pred. No. 6.9e-09;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3576 cgctacgaccagctacagacgacctgggacacacgctc 3614  
|||||

DB 1718 CGCTACGACCAGCTACAGACGACCTGGGACACACGCTC 1756

## RESULT 3

US-08-742-185-101  
Sequence 101, Application US/08742185

Patent No. 6020476

GENERAL INFORMATION:

APPLICANT: Page, David C.

APPLICANT: Reijo, Renee

APPLICANT: Saxena, Richa

APPLICANT: Hawkins, Trevor

APPLICANT: Reeve, Mary Pat

TITLE OF INVENTION: DAZ: A GENE FAMILY ASSOCIATED WITH AZOOSPERMIA

NUMBER OF SEQUENCES: 102

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Militia Drive

CITY: Lexington

STATE: Massachusetts

COUNTRY: US

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/742,185

FILING DATE: 30-OCT-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/690,734

FILING DATE: 31-JUL-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/310,429

FILING DATE: 22-SEP-1994

ATTORNEY/AGENT INFORMATION:

NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227

REFERENCE/DOCKET NUMBER: WHI94-07A2

TELEPHONE: (617) 861-6240

TELEFAX: (617) 861-9540

INFORMATION FOR SEQ ID NO: 101:

SEQUENCE CHARACTERISTICS:

LENGTH: 43795 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-742-185-101

Query Match 0.8%; Score 31; DB 3; Length 43795;

Best Local Similarity 100.0%; Pred. No. 4.4e-05;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2385 gaagaagaagaagaagcagcagaagaagaag 2415  
|||||

DB 38989 GAAGAAGAAGAAGAAGCAGCAGAGAGAGAAG 39019

## RESULT 4

US-08-623-906A-1/C

Sequence 1, Application US/08623906A

Patent No. 5874217

GENERAL INFORMATION:

APPLICANT: Stevenson, Tamara

RESULT 5  
US-08-941-445A-6  
; Sequence 6, Application US/08941445A  
; Patent No. 6107060  
; GENERAL INFORMATION:  
; APPLICANT: Keeling, Peter  
; APPLICANT: Guan, Haining  
; TITLE OF INVENTION: Starch Encapsulation  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.

```

; STREET: 5370 Manhattan Circle
; CITY: Boulder
; STATE: CO
; COUNTRY: US
; ZIP: 80303
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/941,445A
; FILING DATE: 30-SEP-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,855
; FILING DATE: 30-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Winner, Ellen P
; REGISTRATION NUMBER: 28,547
; REFERENCE/DOCKET NUMBER: 89-97
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2542 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 453..2282
;
; US-08-941-445A-6
;
; Query Match 0.7%; Score 25; DB 3; Length 2542;
; Best Local Similarity 100.0%; Pred. No. 0.03;
; Matches 25; Conservative 0; Mismatches 0; Indels 0
;
; QY 3375 tcgttttcgtcgccgcgagatggc 3399
; |||||||
; Db 706 TCGTGTTCGTGGCCGCGAGATGGC 730
;
; RESULT 6
; US-08-860-038-23
; Sequence 23, Application US/08860038
; Patent No. 6287762
; GENERAL INFORMATION:
; APPLICANT: CROUZET, Joel
; APPLICANT: SCHERMAN, Daniel
; APPLICANT: WILS, Pierre
; TITLE OF INVENTION: PURIFICATION OF A TRIPLE HELIX FORMATION
; DATE OF INVENTION: WITH AN IMMOBILIZED OLIGONUCLEOTIDE
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Road, Mailstop 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19426
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

Thu Aug 1 08:29:55 2002

us-09-899-718a-1.oli.rni

Db 8 aggaagaagaagaagaagaag 31

RESULT 8

US-09-370-253-5/c

Sequence 5, Application US/09370253

Patent No. 6165792

GENERAL INFORMATION:

APPLICANT: Allen, Stephen M.

APPLICANT: Sakai, Hajime

APPLICANT: Thorpe, Catherine J.

TITLE OF INVENTION: Amino Acid Transporters

FILE REFERENCE: BB-1200

CURRENT APPLICATION NUMBER: US/09/370,253

CURRENT FILING DATE: 1999-08-09

EARLIER APPLICATION NUMBER: 60/097,222

EARLIER FILING DATE: August 20, 1998

NUMBER OF SEQ ID NOS: 18

SOFTWARE: Microsoft Office 97

SEQ ID NO 5

LENGTH: 1865

TYPE: DNA

ORGANISM: Oryza sativa

US-09-370-253-5

Query Match 0.6%; Score 24; DB 4; Length 1865;

Best Local Similarity 100.0%; Pred. No. 0.09; Mismatches 0; Indels 0; Gaps 0;

Matches 24; Conservative 0;

QY 2377 aggaagaagaagaagaagaag 2400

|||||

Db 50 AGGAAGAAGAAGAAGAAGAAG 27

RESULT 9

US-08-574-959A-8

Sequence 8, Application US/08574959A

Patent No. 5962224

GENERAL INFORMATION:

APPLICANT: Jaekyoon Shin, Insil Joong, Ratna K. Vadlamudi

APPLICANT: and Jack L. Strominger

TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES

TITLE OF INVENTION: AND USES THEREFOR

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD

STREET: 60 State Street, Suite 510

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/574,959A

FILING DATE: 19-DEC-95

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: DFN-008

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 3211 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

APPLICATION NUMBER: US/08/860,038

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 94/15162

FILING DATE: 16-DEC-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO FR95/01468

FILING DATE: 08-NOV-1995

ATTORNEY/AGENT INFORMATION:

NAME: Savitzky Esq., Martin F.

REGISTRATION NUMBER: 29,699

REFERENCE/DOCKET NUMBER: ST94090-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (610) 454-3816

TELEFAX: (610) 454-3808

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 39 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "Oligonucleotide"

US-08-860-038-23

Query Match 0.6%; Score 24; DB 4; Length 39;

Best Local Similarity 100.0%; Pred. No. 0.089;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2377 aggaagaagaagaagaagaag 2400

|||||

Db 8 AGGAAGAAGAAGAAGAAGAAG 31

RESULT 7

US-09-580-923-23

Sequence 23, Application US/09580923

Patent No. 6319672

GENERAL INFORMATION:

APPLICANT: Crouzet, Joel

APPLICANT: Scherman, Daniel

APPLICANT: Wills, Pierre

APPLICANT: Cameron, Beatrice

APPLICANT: Blanche, Francis

TITLE OF INVENTION: PURIFICATION OF A TRIPLE HELIX FORMATION WITH AN

TITLE OF INVENTION: IMMOBILIZED OLIGONUCLEOTIDE

FILE REFERENCE: 03804.0138-01

CURRENT APPLICATION NUMBER: US/09/580,923

CURRENT FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 08/860,038

PRIOR FILING DATE: 1997-06-09

PRIOR APPLICATION NUMBER: PCT/FR95/01468

PRIOR FILING DATE: 1995-11-08

NUMBER OF SEQ ID NOS: 36

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 23

LENGTH: 39

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence:

OTHER INFORMATION: oligonucleotide

US-09-580-923-23

Query Match 0.6%; Score 24; DB 4; Length 39;

Best Local Similarity 100.0%; Pred. No. 0.089;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2377 aggaagaagaagaagaagaag 2400

|||||



; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 439..3157  
US-08-574-959A-8

Query Match 0.6%; Score 24; DB 2; Length 3211;  
Best Local Similarity 100.0%; Pred. No. 0.09;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2377 aggaagaagaagaagaagaag 2400  
|||||  
DB 2423 AGGAAGAAGAAGAAGAAG 2446

RESULT 10  
US-09-357-014-8  
; Sequence 8, Application US/09357014  
; Patent No. 6291645  
; GENERAL INFORMATION:  
; APPLICANT: Jaekyoon Shin, Insil Joong, Ratna K. Vadlamudi  
; and Jack L. Strominger  
; TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES  
; AND USES THEREFOR  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 60 State Street, Suite 510  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/357,014  
; FILING DATE: 19-Jul-1999

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/574,959  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mandragouras, Amy E.  
; REGISTRATION NUMBER: 36,207  
; REFERENCE/DOCKET NUMBER: DFN-008  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3211 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; FEATURE:

NAME/KEY: CDS  
; LOCATION: 439..3157  
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-09-357-014-8

Query Match 0.6%; Score 24; DB 4; Length 3211;  
Best Local Similarity 100.0%; Pred. No. 0.09;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2377 aggaagaagaagaagaagaag 2400  
|||||  
DB 2423 AGGAAGAAGAAGAAGAAG 2446

RESULT 11  
US-08-574-959A-6  
; Sequence 6, Application US/08574959A  
; Patent No. 5962224  
; GENERAL INFORMATION:  
; APPLICANT: Jaekyoon Shin, Insil Joong, Ratna K. Vadlamudi  
; and Jack L. Strominger  
; TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES  
; AND USES THEREFOR  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 60 State Street, Suite 510  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/574,959A  
; FILING DATE: 19-DEC-95

ATTORNEY/AGENT INFORMATION:  
; NAME: Mandragouras, Amy E.  
; REGISTRATION NUMBER: 36,207  
; REFERENCE/DOCKET NUMBER: DFN-008  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3901 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 439..3847  
US-08-574-959A-6

Query Match 0.6%; Score 24; DB 2; Length 3901;  
Best Local Similarity 100.0%; Pred. No. 0.091;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2377 aggaagaagaagaagaagaag 2400  
|||||  
DB 3113 AGGAAGAAGAAGAAGAAG 3136

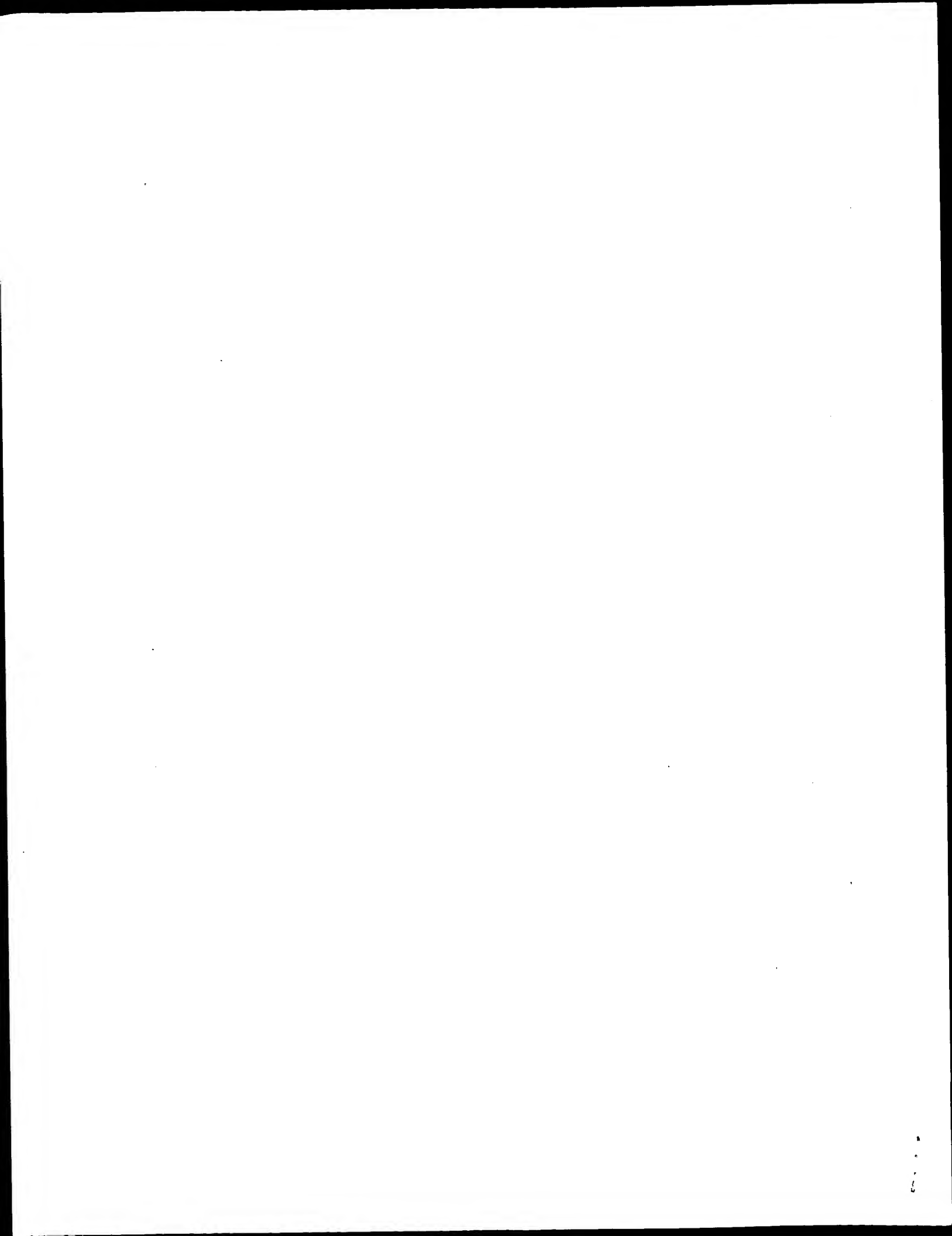
RESULT 12  
US-09-357-014-6  
; Sequence 6, Application US/09357014  
; Patent No. 6291645  
; GENERAL INFORMATION:  
; APPLICANT: Jaekyoon Shin, Insil Joong, Ratna K. Vadlamudi  
; and Jack L. Strominger  
; TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES  
; AND USES THEREFOR  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 60 State Street, Suite 510  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875

COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION NUMBER: US/09/357,014  
;; FILING DATE: 19-Jul-1999  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/574,959  
;; FILING DATE: <Unknown>  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Mandragouras, Amy E.  
;; REGISTRATION NUMBER: 36,207  
;; REFERENCE/DOCKET NUMBER: DFN-008  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617)227-7400  
;; TELEFAX: (617)227-5941  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 3901 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; MOLECULE TYPE: cDNA  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 439..3847  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-09-357-014-6  
  
Query Match 0.6%; Score 24; DB 4; Length 3901;  
Best Local Similarity 100.0%; Pred. No. 0.091;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2377 aggaagaagaagaagaagaag 2400  
|||||  
DB 3113 AGGAAGAAGAAGAAGAAGAAG 3136  
  
RESULT 13  
US-08-688-376-1  
; Sequence 1, Application US/08688376  
; Patent No. 6018039  
; GENERAL INFORMATION:  
; APPLICANT: Satow, Hiroyasu  
; TITLE OF INVENTION: NOVEL PROCESS FOR PRODUCING SUBSTANCES  
; TITLE OF INVENTION: IN MAMMARY GLAND OF TRANSGENIC ANIMAL BY USING MC26 GENE  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DILWORTH & BARRESE  
; STREET: 4350 LaJolla Village Drive, Suite 300  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/688,376  
; FILING DATE: 30-JUL-1996  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pepper, Frederick W.  
; REGISTRATION NUMBER: 31,286  
; REFERENCE/DOCKET NUMBER: 567-3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-546-4410  
; TELEFAX: 619-453-2839

;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 5394 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: CAAT\_signal  
;; LOCATION: 2234..2243  
;; FEATURE:  
;; NAME/KEY: TATA\_signal  
;; LOCATION: 2275..2281  
;; FEATURE:  
;; NAME/KEY: polyA\_signal  
;; LOCATION: 4607..4612  
;; US-08-688-376-1  
  
Query Match 0.6%; Score 24; DB 3; Length 5394;  
Best Local Similarity 100.0%; Pred. No. 0.091;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2377 aggaagaagaagaagaagaag 2400  
|||||  
DB 1437 AGGAAGAAGAAGAAGAAGAAG 1460  
  
RESULT 14  
US-08-553-619B-8/c  
; Sequence 8, Application US/08553619B  
; Patent No. 5919705  
; GENERAL INFORMATION:  
; APPLICANT: Dehaan, Petrus T.  
; TITLE OF INVENTION: Virus Resistant Plants  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5919705artis Crop Protection  
; STREET: 975 California Avenue  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/553,619B  
; FILING DATE: December 1, 1995  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Marcus-Wyner, Lynn  
; REGISTRATION NUMBER: 34,869  
; REFERENCE/DOCKET NUMBER: 137-1082/PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/354-3588  
; TELEFAX: 415/857-1125  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2621 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; ORIGINAL SOURCE:  
; ORGANISM: Chimeric tomato spotted wilt virus S RNA  
; US-08-553-619B-8  
  
Query Match 0.6%; Score 23; DB 2; Length 2621;  
Best Local Similarity 100.0%; Pred. No. 0.27;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 10:43:18 ; Search time 2776.96 seconds  
(without alignments)  
18396.367 Million cell updates/sec

Title: US-09-899-718a-1  
Perfect score: 3785  
Sequence: 1 gtttggttcgtgttttc.....tggaccgcgtgttcgtcgac 3785

Scoring table:

OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID       | Description        |
|------------|-------|-------------|--------|----|----------|--------------------|
| 1          | 322   | 8.5         | 527    | 9  | AW448831 | AW448831 BRY_1580  |
| 2          | 52    | 1.4         | 462    | 9  | AL508396 | AL508396 BRY_1580  |
| 3          | 52    | 1.4         | 513    | 9  | AL508933 | AL508933 AL508933  |
| 4          | 52    | 1.4         | 558    | 10 | BE602243 | BE602243 HVSMEH009 |
| 5          | 52    | 1.4         | 638    | 9  | AL508959 | AL508959 AL508959  |
| 6          | 52    | 1.4         | 700    | 9  | AL508567 | AL508567 AL508567  |
| 7          | 52    | 1.4         | 700    | 9  | AL508163 | AL508163 AL508163  |
| 8          | 51    | 1.3         | 579    | 10 | BE402419 | BE402419 CSB007G05 |
| 9          | 51    | 1.3         | 639    | 10 | BE414303 | BE414303 SCU008.CO |
| 10         | 44    | 1.2         | 421    | 12 | AZ113646 | AZ113646 RPCI-23-4 |
| 11         | 41    | 1.1         | 494    | 9  | AW448811 | AW448811 BRY_1487  |
| 12         | 40    | 1.1         | 444    | 10 | BM368238 | BM368238 EBE01_SQ  |
| 13         | 37    | 1.0         | 165    | 12 | AZ733073 | AZ733073 RPCI-24-1 |
| 14         | 37    | 1.0         | 260    | 12 | AZ462781 | AZ462781 IM0271G17 |
| 15         | 37    | 1.0         | 380    | 12 | AZ368386 | AZ368386 IM0118M19 |
| 16         | 37    | 1.0         | 390    | 12 | AZ063578 | AZ063578 RPCI-23-4 |
| 17         | 37    | 1.0         | 751    | 10 | BE373417 | BE373417 601225078 |

|      |    |     |      |    |          |                    |
|------|----|-----|------|----|----------|--------------------|
| c 18 | 36 | 1.0 | 179  | 12 | BH122005 | BH122005 RPCI-24-3 |
| c 19 | 35 | 0.9 | 757  | 12 | A0740608 | A0740608 HS_5508_A |
| c 20 | 33 | 0.9 | 446  | 10 | BJ163167 | BJ163167 BJ163167  |
| c 21 | 33 | 0.9 | 514  | 12 | BH322444 | BH322444 CH230-197 |
| c 22 | 32 | 0.8 | 305  | 12 | A2358393 | A2358393 IM0100L13 |
| c 23 | 32 | 0.8 | 324  | 12 | A2454883 | A2454883 IM0257G06 |
| c 24 | 31 | 0.8 | 166  | 12 | A2339932 | A2339932 IM0071E10 |
| c 25 | 31 | 0.8 | 258  | 9  | BS544345 | BS544345 BS544345  |
| c 26 | 31 | 0.8 | 333  | 10 | BG075200 | BG075200 H3144502- |
| c 27 | 31 | 0.8 | 369  | 10 | BF893280 | BF893280 QV3-MT012 |
| c 28 | 31 | 0.8 | 455  | 10 | T44352   | T44352 7615 Lambda |
| c 29 | 31 | 0.8 | 585  | 10 | BG908116 | BG908116 Tatr1165F |
| c 30 | 31 | 0.8 | 952  | 10 | BF101949 | BF101949 601752915 |
| c 31 | 30 | 0.8 | 279  | 9  | AJ281579 | AJ281579 4A3A-P7A8 |
| c 32 | 30 | 0.8 | 452  | 9  | AW596355 | AW596355 SJ02d02.Y |
| c 33 | 30 | 0.8 | 465  | 10 | T46573   | T46573 9836 Lambda |
| c 34 | 30 | 0.8 | 518  | 12 | AZ771492 | AZ771492 IM0573L10 |
| c 35 | 30 | 0.8 | 848  | 12 | CNS05JRH | AL340298 Tetraodon |
| c 36 | 30 | 0.8 | 907  | 12 | CNS04OUG | AL300481 Tetraodon |
| c 37 | 30 | 0.8 | 922  | 12 | CNS032JN | AL24924 Tetraodon  |
| c 38 | 30 | 0.8 | 968  | 12 | CNS05T11 | AL352558 Tetraodon |
| c 39 | 30 | 0.8 | 1020 | 12 | CNS03OJH | AL253430 Tetraodon |
| c 40 | 30 | 0.8 | 1050 | 12 | CNS059X1 | AL327790 Tetraodon |
| c 41 | 30 | 0.8 | 1062 | 12 | CNS059RS | AL327601 Tetraodon |
| c 42 | 30 | 0.8 | 1073 | 12 | CNS05G60 | AL335913 Tetraodon |
| c 43 | 30 | 0.8 | 1101 | 12 | CNS05751 | AL324207 Tetraodon |
| c 44 | 30 | 0.8 | 1101 | 12 | CNS05APL | AL328818 Tetraodon |
| c 45 | 29 | 0.8 | 94   | 12 | AZ347674 | AZ347674 IM0083C22 |

## ALIGNMENTS

### RESULT 1

AW448831

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AW448831 BRY\_1580 BRY Triticum aestivum CDNA clone P56-1P, mRNA sequence. EST 03-JAN-2001

AW448831.1 GI:12019366

EST.

bread wheat.

Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae

; Triticeae; Triticum.

1 (bases 1 to 527)

Clarke,B.C., Hobbs,M. and Appels,R.

Genes active in developing wheat endosperm

Unpublished (2000)

Contact: Bryan Clarke

Division of Plant Industry

C.S.I.R.O.

GPO Box 1600, Canberra, ACT, Australia

Tel: 61 2 6246 5054

Fax: 61 2 6246 5000

Email: bryan@dpi.csiro.au

Location/Qualifiers

1. 527

/organism="Triticum aestivum"

/cultivar="Wyuna"

/db\_xref="taxon:4565"

/clone="p56-1p"

/clone\_lib="BRY"

/cell\_type="endosperm"

115 a 176 c 163 g 73 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 322; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3099 cagtcctctgtctgcaggtacccaccctgcgcgcacatgcgcgcgtctgtcagct 3158

|                       |         |                                                |                                                        |    |            |      |        |                 |      |    |
|-----------------------|---------|------------------------------------------------|--------------------------------------------------------|----|------------|------|--------|-----------------|------|----|
|                       | Matches | 52;                                            | Conservative                                           | 0; | Mismatches | 0;   | Indels | 0;              | Gaps | 0; |
| QY                    | 3363    | gagcattgaactgttcttcgcggcccgatgaggcccttgacgaac  | 3414                                                   |    |            |      |        |                 |      |    |
| Db                    | 174     | GGCGATGAACCTGTTGTTCGCGCCGAGATGGCGCCTGGAGCAAGAC | 225                                                    |    |            |      |        |                 |      |    |
| RESULT                | 3       |                                                |                                                        |    |            |      |        |                 |      |    |
| AL508933              |         | AL508933                                       | Hordeum vulgare Barke developing caryopsis (3.-15.DAP) |    | 513 bp     | mRNA | linear | EST 04-JAN-2001 |      |    |
| LOCUS                 |         | AL508933                                       | Hordeum vulgare cdna clone H10D16V 5', mRNA sequence.  |    |            |      |        |                 |      |    |
| DEFINITION            |         | AL508933                                       | Hordeum vulgare EST.                                   |    |            |      |        |                 |      |    |
| ACCESSION             |         | AL508933                                       | AL508933.1 GI:12035436                                 |    |            |      |        |                 |      |    |
| VERSION               |         |                                                |                                                        |    |            |      |        |                 |      |    |
| KEYWORDS              |         |                                                |                                                        |    |            |      |        |                 |      |    |
| SOURCE                |         |                                                |                                                        |    |            |      |        |                 |      |    |
| ORGANISM              |         |                                                |                                                        |    |            |      |        |                 |      |    |
| REFERENCE             |         |                                                |                                                        |    |            |      |        |                 |      |    |
| AUTHORS               |         |                                                |                                                        |    |            |      |        |                 |      |    |
| TITLE                 |         |                                                |                                                        |    |            |      |        |                 |      |    |
| JOURNAL               |         |                                                |                                                        |    |            |      |        |                 |      |    |
| COMMENT               |         |                                                |                                                        |    |            |      |        |                 |      |    |
| FEATURES              |         |                                                |                                                        |    |            |      |        |                 |      |    |
| source                |         |                                                |                                                        |    |            |      |        |                 |      |    |
| BASE COUNT            |         |                                                |                                                        |    |            |      |        |                 |      |    |
| ORIGIN                |         |                                                |                                                        |    |            |      |        |                 |      |    |
| Query Match           |         |                                                |                                                        |    |            |      |        |                 |      |    |
| Best Local Similarity |         |                                                |                                                        |    |            |      |        |                 |      |    |
| Matches               |         |                                                |                                                        |    |            |      |        |                 |      |    |
| QY                    | 3363    | gagcattgaactgttcttcgcggcccgatgaggcccttgacgaac  | 3414                                                   |    |            |      |        |                 |      |    |
| Db                    | 452     | GGCGATGAACCTGTTGTTCGCGCCGAGATGGCGCCTGGAGCAAGAC | 503                                                    |    |            |      |        |                 |      |    |
| RESULT                | 4       |                                                |                                                        |    |            |      |        |                 |      |    |
| BE602243              |         | BE602243                                       | Hordeum vulgare EST.                                   |    |            |      |        |                 |      |    |
| LOCUS                 |         | BE602243                                       | Hordeum vulgare EST.                                   |    |            |      |        |                 |      |    |
| DEFINITION            |         | BE602243                                       | Hordeum vulgare EST.                                   |    |            |      |        |                 |      |    |
| ACCESSION             |         | BE602243                                       | BE602243.2 GI:13189982                                 |    |            |      |        |                 |      |    |
| VERSION               |         |                                                |                                                        |    |            |      |        |                 |      |    |
| KEYWORDS              |         |                                                |                                                        |    |            |      |        |                 |      |    |
| SOURCE                |         |                                                |                                                        |    |            |      |        |                 |      |    |
| BASE COUNT            |         |                                                |                                                        |    |            |      |        |                 |      |    |
| ORIGIN                |         |                                                |                                                        |    |            |      |        |                 |      |    |
| Query Match           |         |                                                |                                                        |    |            |      |        |                 |      |    |
| Best Local Similarity |         |                                                |                                                        |    |            |      |        |                 |      |    |
| Matches               |         |                                                |                                                        |    |            |      |        |                 |      |    |
| QY                    | 462     | gagcattgaactgttcttcgcggcccgatgaggcccttgacgaac  | 462                                                    |    |            |      |        |                 |      |    |
| Db                    | 72 t    | GGCGATGAACCTGTTGTTCGCGCCGAGATGGCGCCTGGAGCAAGAC | 72 t                                                   |    |            |      |        |                 |      |    |
| RESULT                | 7       |                                                |                                                        |    |            |      |        |                 |      |    |
| AL506396              |         | AL506396                                       | Hordeum vulgare Barke developing caryopsis (3.-15.DAP) |    | 462 bp     | mRNA | linear | EST 04-JAN-2001 |      |    |
| LOCUS                 |         | AL506396                                       | Hordeum vulgare cdna clone HY02P18T 5', mRNA sequence. |    |            |      |        |                 |      |    |
| DEFINITION            |         | AL506396                                       | Hordeum vulgare EST.                                   |    |            |      |        |                 |      |    |
| ACCESSION             |         | AL506396                                       | AL506396.1 GI:12032611                                 |    |            |      |        |                 |      |    |
| VERSION               |         |                                                |                                                        |    |            |      |        |                 |      |    |
| KEYWORDS              |         |                                                |                                                        |    |            |      |        |                 |      |    |
| SOURCE                |         |                                                |                                                        |    |            |      |        |                 |      |    |
| ORGANISM              |         |                                                |                                                        |    |            |      |        |                 |      |    |
| REFERENCE             |         |                                                |                                                        |    |            |      |        |                 |      |    |
| AUTHORS               |         |                                                |                                                        |    |            |      |        |                 |      |    |
| TITLE                 |         |                                                |                                                        |    |            |      |        |                 |      |    |
| JOURNAL               |         |                                                |                                                        |    |            |      |        |                 |      |    |
| COMMENT               |         |                                                |                                                        |    |            |      |        |                 |      |    |
| FEATURES              |         |                                                |                                                        |    |            |      |        |                 |      |    |
| source                |         |                                                |                                                        |    |            |      |        |                 |      |    |
| BASE COUNT            |         |                                                |                                                        |    |            |      |        |                 |      |    |
| ORIGIN                |         |                                                |                                                        |    |            |      |        |                 |      |    |
| Query Match           |         |                                                |                                                        |    |            |      |        |                 |      |    |
| Best Local Similarity |         |                                                |                                                        |    |            |      |        |                 |      |    |
| Matches               |         |                                                |                                                        |    |            |      |        |                 |      |    |
| QY                    | 462     | gagcattgaactgttcttcgcggcccgatgaggcccttgacgaac  | 462                                                    |    |            |      |        |                 |      |    |
| Db                    | 72 t    | GGCGATGAACCTGTTGTTCGCGCCGAGATGGCGCCTGGAGCAAGAC | 72 t                                                   |    |            |      |        |                 |      |    |
| RESULT                | 7       |                                                |                                                        |    |            |      |        |                 |      |    |
| AL506396              |         | AL506396                                       | Hordeum vulgare Barke developing caryopsis (3.-15.DAP) |    | 462 bp     | mRNA | linear | EST 04-JAN-2001 |      |    |
| LOCUS                 |         | AL506396                                       | Hordeum vulgare cdna clone HY02P18T 5', mRNA sequence. | </ |            |      |        |                 |      |    |

ORGANISM Hordeum vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae  
; Triticeae; Hordeum.  
1 (bases 1 to 558)  
Wing,R., Close,T.J., Kleinbols,A., Wise,R., Begum,D., Frisch,D., Yu  
Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton  
R.D., Close,S.J., Oates,R. and Main,D.  
Development of a genetically and physically anchored EST resource  
for barley genomics: Morex 5-45 DAP spike cDNA library  
Unpublished (2001)  
On Aug 21, 2000 this sequence version replaced gi:9859804.  
Contact: Wing RA  
Clemson University  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Total hq bases = 215  
Seq primer: AATTAACTCTACTAAAGGG  
High quality sequence start: 16  
High quality sequence stop: 550.

FEATURES  
source  
1..558  
Location/Qualifiers  
/organism="Hordeum vulgare"  
/cultivar="Morex"  
/db\_xref="taxon:4513"  
/clone="HVSHEH0098D06f"  
/HVCDA0009 (5 to 45 DAP)  
/issue\_type="5-45 DAP Spike"  
/lab\_host="SOLR"  
/note="Vector: lambdaZAP; Site\_1: EcoRI; Site\_2: XhoI;  
Plants were grown in the greenhouse at the University of  
California, Riverside (Fenton, SJ Close, TJ Close). Whole  
spikes with awns trimmed were collected at 5, 10, 15, 20,  
30 and 45 DAP (Fenton). Total RNA was prepared from each  
pool, equal quantities of all six RNA pools were combined,  
poly(A) RNA was purified from the mixture, one primary  
unamplified cDNA library was made, and 1 million pfu were  
in vivo excised to give pBluescript SK(-) cDNA phagemids  
(Choi) in the TJ Close lab at the University of California,  
Riverside. Phagemids were plated and picked at the Clemson  
University Genomics Institute (CUGI) (Begum, Palmer,  
Frisch, Atkins and Wing). Plasmid DNA preparations, DNA  
sequencing and sequence analysis were performed at CUGI  
(Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).  
The sequence has been trimmed to remove vector sequence  
and contains a minimum of 100 bases of phred value 20 or  
above. For more details on library preparation and  
sequence analysis see  
http://www.genome.clemson.edu/projects/barley. To order  
this clone see http://www.genome.clemson.edu/orders/Also  
see Close TJ, Wing R, Kleinbols A, Wise R (2001)  
Genetically and physically anchored EST resources for  
barley genomics. Barley Genetics Newsletter 31:29-30.  
(http://wheat.pw.usda.gov/gvpages/bgn/31/cover.html)."

BASE COUNT 102 a 209 c 188 g 78 t 1 others  
ORIGIN

Query Match 1.4%; Score 52; DB 10; Length 558;  
Best Local Similarity 100.0%; Pred. No. 3.3e-13;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3363 gcggcatgaacctgtgttcgcccgcgagatggccctggagcaagac 3414  
|||||  
Db 440 GCGCATGAACCTCGTGTTCGCGCGGAGATGGCGCCCTGGAGCAAGAC 491  
|||||

RESULT 5  
AL508959

LOCUS AL508959 638 bp mRNA linear EST 04-JAN-2001  
DEFINITION Hordeum vulgare Barke developing caryopsis (3.-15.DAP)  
Hordeum vulgare cDNA clone HY10E20V 5', mRNA sequence.  
ACCESSION AL508959  
VERSION AL508959.1 GI:12035462  
KEYWORDS EST.  
SOURCE Hordeum vulgare  
ORGANISM Hordeum vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae  
; Triticeae; Hordeum.  
1 (bases 1 to 638)  
Michalek,W., Weschke,W., Pleissner,K.-P. and Graner,A.  
EST sequencing and analysis in barley  
Unpublished (2000)  
Contact: Michalek W  
Institute for Plant Genetics and Crop Plant Research  
Corrensstr.3, D-06466 Gatersleben, Germany  
Email: michalek@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de  
Seq primer: T3 primer for 5'end.

FEATURES  
source  
1..638  
Location/Qualifiers  
/organism="Hordeum vulgare"  
/cultivar="Barke"  
/db\_xref="taxon:4513"  
/clone="HY10E20V"  
/clone\_lib="Hordeum vulgare Barke developing caryopsis  
(3.-15.DAP)"  
/tissue\_type="developing caryopsis (3.-15.DAP)"  
/lab\_host="XLOLR"  
/note="Vector: plasmid pBK-CMV; Site\_1: EcoRI; Site\_2:  
XhoI; mRNA was made from developing caryopsis (3.-15.DAP)  
of spring barley variety 'Barke', a high quality malting  
variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI  
(3'-end of cDNA). NOTE: Due to a cloning artefact caused  
by the kit, in most cases the EcoRI site is NOT present,  
as well as the EcoRI adapter. Average insert size is 1 kb  
sequence trimming: Vector sequences and sequence ends were  
trimmed from the 5'-and 3'-end until a 50 bp window  
contains less than two ambiguities. The maximum length was  
set to 700 bp"

BASE COUNT 119 a 235 c 190 g 93 t 1 others  
ORIGIN

Query Match 1.4%; Score 52; DB 9; Length 638;  
Best Local Similarity 100.0%; Pred. No. 3.2e-13;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3363 gcggcatgaacctgtgttcgcccgcgagatggccctggagcaagac 3414  
|||||  
Db 447 GCGCATGAACCTCGTGTTCGCGCGGAGATGGCGCCCTGGAGCAAGAC 498  
|||||

RESULT 6  
AL506567

LOCUS AL506567 700 bp mRNA linear EST 04-JAN-2001  
DEFINITION Hordeum vulgare Barke developing caryopsis (3.-15.DAP)  
Hordeum vulgare cDNA clone HY03121T 5', mRNA sequence.  
ACCESSION AL506567  
VERSION AL506567.1 GI:12032782  
KEYWORDS EST.  
SOURCE Hordeum vulgare  
ORGANISM Hordeum vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae  
; Triticeae; Hordeum.  
1 (bases 1 to 700)  
Michalek,W., Weschke,W., Pleissner,K.-P. and Graner,A.  
EST sequencing and analysis in barley  
Unpublished (2000)  
Contact: Michalek W  
Institute for Plant Genetics and Crop Plant Research



Corrensstr.3, D-06466 Gatersleben, Germany  
 Email: michalek@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de  
 Seq primer: T3 primer for 5'end.

#### FEATURES

Location/Qualifiers  
 1..700  
 /organism="Hordeum vulgare"  
 /cultivar="Barke"  
 /db\_xref="taxon:4513"  
 /clone="HY03121T"  
 /clone\_lib="Hordeum vulgare Barke developing caryopsis  
 (3.-15.DAP)"  
 /tissue\_type="developing caryopsis (3.-15.DAP)"  
 /lab\_host="XL0LR"  
 /note="Vector: Plasmid pBK-CMV; Site1: EcoRI; Site2:  
 XhoI; mRNA was made from developing caryopsis (3.-15.DAP)  
 of spring barley variety 'Barke', a high quality malting  
 variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI  
 (3'-end of cDNA). NOTE: Due to a cloning artefact caused  
 by the kit, in most cases the EcoRI site is NOT present,  
 as well as the EcoRI adapter. Average insert size is 1 kb  
 Sequence trimming: Vector sequences and sequence ends were  
 trimmed from the 5'-and 3'-end until a 50 bp window  
 contains less than two ambiguities. The maximum length was  
 set to 700 bp"

BASE COUNT 134 a 229 c 218 g 117 t 2 others  
 ORIGIN

Query Match 1.4%; Score 52; DB 9; Length 700;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-13;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3363 gcggcatgaacctgttcgtcgccgagatggcgccctggagcaagac 3414  
 |||||  
 Db 198 GCGGCATGAACCTGCTTCGTGCGCGGAGATGGCGCCCTGGAGCAAGAC 249

#### RESULT

AL508163 700 bp mRNA linear EST 04-JAN-2001  
 LOCUS  
 DEFINITION  
 Hordeum vulgare Barke developing caryopsis (3.-15.DAP)  
 Hordeum vulgare cDNA clone HY07P10V 5', mRNA sequence.

ACCESSION  
 AL508163  
 VERSION  
 AL508163.1 GI:12034378  
 EST.  
 KEYWORDS  
 SOURCE

ORGANISM  
 Hordeum vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae  
 ; Triticeae; Hordeum.

REFERENCE  
 1 (bases 1 to 700)  
 AUTHORS  
 Michalek,W., Weschke,W., Pleissner,K.-P. and Graner,A.  
 TITLE  
 EST sequencing and analysis in barley

JOURNAL  
 Unpublished (2000)  
 COMMENT  
 Contact: Michalek W  
 Institute for Plant Genetics and Crop Plant Research  
 Corrensstr.3, D-06466 Gatersleben, Germany  
 Email: michalek@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de  
 Seq primer: T3 primer for 5'end.

#### FEATURES

Location/Qualifiers  
 1..700  
 /organism="Hordeum vulgare"  
 /cultivar="Barke"  
 /db\_xref="taxon:4513"  
 /clone="HY07P10V"  
 /clone\_lib="Hordeum vulgare Barke developing caryopsis  
 (3.-15.DAP)"  
 /tissue\_type="developing caryopsis (3.-15.DAP)"  
 /lab\_host="XL0LR"  
 /note="Vector: Plasmid pBK-CMV; Site1: EcoRI; Site2:  
 XhoI; mRNA was made from developing caryopsis (3.-15.DAP)  
 of spring barley variety 'Barke', a high quality malting  
 variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI

(3'-end of cDNA). NOTE: Due to a cloning artefact caused  
 by the kit, in most cases the EcoRI site is NOT present,  
 as well as the EcoRI adapter. Average insert size is 1 kb  
 Sequence trimming: Vector sequences and sequence ends were  
 trimmed from the 5'-and 3'-end until a 50 bp window  
 contains less than two ambiguities. The maximum length was  
 set to 700 bp"

BASE COUNT 133 a 245 c 213 g 107 t 2 others  
 ORIGIN

Query Match 1.4%; Score 52; DB 9; Length 700;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-13;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3363 gcggcatgaacctgttcgtcgccgagatggcgccctggagcaagac 3414  
 |||||  
 Db 421 GCGGCATGAACCTGCTTCGTGCGCGGAGATGGCGCCCTGGAGCAAGAC 472

#### RESULT

BE402419 579 bp mRNA linear EST 21-JUL-2000  
 LOCUS  
 DEFINITION  
 CSB007G05F990908 ITEC CSB Wheat Endosperm Library Triticum aestivum  
 cDNA clone CSB007G05, mRNA sequence.

ACCESSION  
 BE402419  
 VERSION  
 BE402419.1 GI:9361887  
 EST.  
 KEYWORDS  
 SOURCE

ORGANISM  
 bread wheat.  
 Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae  
 ; Triticeae; Triticum.

REFERENCE  
 1 (bases 1 to 579)  
 AUTHORS

Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier  
 ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,  
 Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,  
 Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,  
 Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,  
 Sorrells,M., Warburton,M. and Wenzel,G.  
 International Triticeae EST Cooperative (ITEC): Production of  
 Expressed Sequence Tags for Species of the Triticeae

#### JOURNAL

Unpublished (2000)  
 COMMENT  
 Contact: Appels R

Div. of Plant Industry, CSIRO  
 Canberra ACT 2601 AUSTRALIA  
 Tel: 61 62 465496  
 Fax: 61 62 465000

Email: rudispi.csiro.au  
 International Triticeae EST Cooperative (ITEC)  
 http://wheat.pw.usda.gov/genome.

#### FEATURES

Location/Qualifiers  
 1..579  
 /organism="Triticum aestivum"  
 /cultivar="Wyuna"  
 /db\_xref="taxon:4565"  
 /clone="CSB007G05"  
 /clone\_lib="ITEC CSB Wheat Endosperm Library"  
 /tissue\_type="endosperm"  
 /dev\_stage="8-12 days post anthesis"  
 /lab\_host="Escherichia coli SOLR"

/note="Vector: Lambda Zap/Bluescript; Site1: XhoI;  
 Site2: EcoRI; Plants grown in phytotron with 18C/13C  
 (day/night) 16 hour light. M13 Reverse sequencing primer  
 used, 1.0 kbp average insert size."  
 BASE COUNT 109 a 207 c 189 g 72 t 2 others  
 ORIGIN

Query Match 1.3%; Score 51; DB 10; Length 579;  
 Best Local Similarity 100.0%; Pred. No. 9.7e-13;  
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



source

1..494  
 /organism="Triticum aestivum"  
 /cultivar="Wyuna"  
 /db\_xref="taxon:4565"  
 /clone="P51-2A"  
 /clone\_lib="BRY"  
 /cell\_type="endosperm"  
 98 a 170 c 160 g 66 t

BASE COUNT  
 ORIGIN

Query Match 1.1%; Score 41; DB 9; Length 494;  
 Best Local Similarity 100.0%; Pred. No. 5.5e-08;  
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3322 ccggcggtgctctccatgtgtgctgcgcacgagcgagc 3362

|||||  
 Db 391 CCGCGGTCCTCTCCATGTTGTGTCGCCACGCGACGC 431

RESULT 12

BM368238

LOCUS

DEFINITION EBed01\_SQ002\_J02\_R IGF Barley EBed01 library Hordeum vulgare cDNA

ACCESSION BM368238

VERSION

KEYWORDS EST.

SOURCE

ORGANISM

Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae

; Triticeae; Hordeum.

1 (bases 1 to 444)

Authors Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,

Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.

Title Development of Barley Transcriptome Resources

Journal Unpublished (2001)

Comment Contact: Waugh R

Unit of Genomics

Scottish Crop Research Institute

Invergowrie, Dundee, DD2 5DA, Scotland, UK

Tel: 00 44 1382 562731

Fax: 00 44 1382 562426

Email: rwaugh@scri.sari.ac.uk

All sequence has a Phred quality score of 20 or over

Seq primer: M13 reverse.

Class: BAC ends.

Location/Qualifiers

1..444

/organism="Hordeum vulgare"

/cultivar="Optic"

/db\_xref="taxon:4513"

/clone="EBed01\_SQ002\_J02"

/clone\_lib="IGF Barley EBed01 library"

/tissue\_type="Endosperm"

/dev\_stage="6 days post anthesis"

/lab\_host="PH108"

/note="Vector: pSPORT1; Site\_1: Sal I; Site\_2: Not I;

Non-normalised library, directionally cloned into pSPORT1.

Derived from endosperm tissue dissected from developing

grains (6 days post anthesis) in glasshouse grown barley

plants. Developed as part of the barley transcriptome

resources of BSRG/SEERAD funded cereal IGF (Investigating

Gene Function) project."

BASE COUNT 90 a 160 c 131 g 63 t

ORIGIN

Query Match

Best Local Similarity 1.1%; Score 40; DB 10; Length 444;  
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3189 ggcgcacgacagattccggcgtccaggttttcaggcct 3228

Db 244 GCGTCACGACAGATTCGCGCGTCCAGGTTTTCAGGCGCT 283

RESULT 13

AZ733073/c

LOCUS

DEFINITION

DNA sequence.

165 bp DNA linear GSS 25-JAN-2001

RPIC-24-118F9-TV RPIC-24 Mus musculus genomic clone RPIC-24-118F9,

ACCESSION AZ733073

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 165)

Authors Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,

Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E.,

Russell, D., de Jong, P. and Fraser, C.M.

Title Mouse BAC End Sequences from Library RPIC-24

Journal Unpublished (1999)

Comment Other\_GSSs: RPIC-24-118F9.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPIC-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end

page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)

Plate: 118 Row: F Column: 9

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1..165

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPIC-24-118F9"

/clone\_lib="RPIC-24"

/sex="Male"

/cell\_type="Spleen/Brain"

/note="Vector: pTARBAC1; Site\_1: BamHI; Site\_2: BamHI;

RPIC-24 Mouse BAC Library produced by Pieter de Jong. The

library was cloned in the pTARBAC1 cloning vector at the

BamHI sites using MboI partially digested male C57BL/6J

DNA."

BASE COUNT 19 a 44 c 12 g 90 t

ORIGIN

Query Match

Best Local Similarity 1.0%; Score 37; DB 12; Length 165;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2379 gaagaagaagaagaagaagaagaagaagaagaagaag 2415

|||||

Db 146 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 110

RESULT 14

AZ462781/c

LOCUS

DEFINITION

clone UGCI1M0271G17 F, DNA sequence.

Accession AZ462781

Version

Keywords

AZ462781.1 GI:10620906

GSS.

AZ462781 260 bp DNA linear GSS 04-OCT-2000  
 IM0271G17F Mouse 10Kb plasmid UGCI1M library Mus musculus genomic  
 clone UGCI1M0271G17 F, DNA sequence.

```

SOURCE      house mouse.
ORGANISM    Mus musculus
REFERENCE   1 (bases 1 to 260)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
TITLE       Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL     Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0271 row: G column: 17
            Seq primer: CGTGTAAACGACGCCAGT
            Class: plasmid ends
            High quality sequence stop: 260.
FEATURES    source
            1..260
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUGC1M0271G17"
                /clone_lib="Mouse 10kb plasmid UUGC1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                /note="Vector: PWD42nv; Purified genomic DNA from M.
            musculus C57BL/6J (male) was obtained from the Jackson
            Laboratory Mouse DNA Resource
            (http://www.jax.org/resources/documents/dnares/). The DNA
            was hydrodynamically sheared by repeated passage through a
            0.005 inch orifice at constant velocity. The sheared DNA
            was blunt end-repaired with T4 DNA polymerase and T4
            polynucleotide kinase. Adaptor oligonucleotides were
            ligated to the blunt ends in high molar excess. The
            adaptor DNA was purified and size-selected for a 9.5 to
            10.5 kb range using preparative agarose gel
            electrophoresis. Vector DNA was prepared from a derivative
            of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
            inducible derivative of plasmid R1. The vector was ligated
            with adaptors complementary to the insert adaptors and
            purified. The sheared, adaptor mouse DNA was annealed to
            adaptor vector DNA, and transformed into
            chemically-competent E. coli XL10-Gold (Stratagene) cells
            and selected for ampicillin resistance."
BASE COUNT  49 a 75 c 30 g 106 t
ORIGIN
Query Match      1.0%; Score 37; DB 12; Length 260;
Best Local Similarity 100.0%; Pred. No. 4.9e-06;
Matches 37; Conservativity 0; Mismatches 0; Indels 0; Gaps 0;
QY 2379 gaagaagaagaagaagaagaagaagaagaagaag 2415
      |||||||
Db 180 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 144
      |||||||

RESULT 15
LOCUS   AZ368386/c
DEFINITION  lm0118m19f Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0118M19 F, DNA sequence.
ACCESSION  AZ368386
VERSION    AZ368386.1 GI:10482086

```

```

KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
REFERENCE   1 (bases 1 to 380)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
TITLE       Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL     Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0118 row: M column: 19
            Seq primer: CGTGTAAACGACGCCAGT
            Class: plasmid ends
            High quality sequence stop: 380.
FEATURES    source
            1..380
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUGC1M0118M19"
                /clone_lib="Mouse 10kb plasmid UUGC1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                /note="Vector: PWD42nv; Purified genomic DNA from M.
            musculus C57BL/6J (male) was obtained from the Jackson
            Laboratory Mouse DNA Resource
            (http://www.jax.org/resources/documents/dnares/). The DNA
            was hydrodynamically sheared by repeated passage through a
            0.005 inch orifice at constant velocity. The sheared DNA
            was blunt end-repaired with T4 DNA polymerase and T4
            polynucleotide kinase. Adaptor oligonucleotides were
            ligated to the blunt ends in high molar excess. The
            adaptor DNA was purified and size-selected for a 9.5 to
            10.5 kb range using preparative agarose gel
            electrophoresis. Vector DNA was prepared from a derivative
            of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
            inducible derivative of plasmid R1. The vector was ligated
            with adaptors complementary to the insert adaptors and
            purified. The sheared, adaptor mouse DNA was annealed to
            adaptor vector DNA, and transformed into
            chemically-competent E. coli XL10-Gold (Stratagene) cells
            and selected for ampicillin resistance."
BASE COUNT  51 a 116 c 18 g 195 t
ORIGIN
Query Match      1.0%; Score 37; DB 12; Length 380;
Best Local Similarity 100.0%; Pred. No. 4.5e-06;
Matches 37; Conservativity 0; Mismatches 0; Indels 0; Gaps 0;
QY 2379 gaagaagaagaagaagaagaagaagaagaagaag 2415
      |||||||
Db 373 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 337
      |||||||

Search completed: July 31, 2002, 14:54:43
Job time: 15085 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:02:43 ; Search time 389.91 Seconds

(without alignments)

16665.721 Million cell updates/sec

Title: US-09-899-718A-1

Perfect score: 3785

Sequence: 1 gtttggttcgtgttttc.....tgaccggtgttcgtgcac 3785

Scoring table:

OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

N\_Geneseq\_032802.\*

- 1: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1980.DAT.\*
- 2: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1981.DAT.\*
- 3: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1982.DAT.\*
- 4: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1983.DAT.\*
- 5: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1984.DAT.\*
- 6: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1985.DAT.\*
- 7: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1986.DAT.\*
- 8: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1987.DAT.\*
- 9: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1988.DAT.\*
- 10: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1989.DAT.\*
- 11: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1990.DAT.\*
- 12: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1991.DAT.\*
- 13: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1992.DAT.\*
- 14: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1993.DAT.\*
- 15: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1994.DAT.\*
- 16: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1995.DAT.\*
- 17: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1996.DAT.\*
- 18: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1997.DAT.\*
- 19: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1998.DAT.\*
- 20: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1999.DAT.\*
- 21: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA2000.DAT.\*
- 22: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA2001A.DAT.\*
- 23: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA2001B.DAT.\*
- 24: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID          | Description        |
|------------|-------|-------------|--------|-------------|--------------------|
| 1          | 39    | 1.0         | 333    | 21 AAC74569 | Human ORFX ORF124  |
| 2          | 39    | 1.0         | 1915   | 19 AAX60319 | DNA sequence of th |
| 3          | 39    | 1.0         | 2267   | 18 AAX63355 | Granule bound star |
| 4          | 39    | 1.0         | 4800   | 19 AAV29752 | Zea mays waxy gene |
| 5          | 37    | 1.0         | 276    | 22 ABA68266 | Human foetal liver |
| 6          | 37    | 1.0         | 276    | 22 ABA35275 | Probe #13741 for g |
| 7          | 37    | 1.0         | 276    | 22 AAK42405 | Human bone marrow  |
| 8          | 37    | 1.0         | 276    | 22 AAI23164 | Probe #13097 for q |
| 9          | 37    | 1.0         | 460    | 23 AAS71148 | DNA encoding novel |

|    |    |     |       |             |                    |
|----|----|-----|-------|-------------|--------------------|
| 10 | 37 | 1.0 | 508   | 23 AAS69548 | DNA encoding novel |
| 11 | 37 | 1.0 | 771   | 23 AAS90697 | DNA encoding novel |
| 12 | 37 | 1.0 | 2187  | 23 AAS75468 | DNA encoding novel |
| 13 | 31 | 0.8 | 711   | 23 AAS90716 | DNA encoding novel |
| 14 | 31 | 0.8 | 43795 | 21 AAZ92583 | Human DAZ genomic  |
| 15 | 29 | 0.8 | 509   | 21 AAC29802 | Arabidopsis thalia |
| 16 | 29 | 0.8 | 512   | 21 AAC39876 | Arabidopsis thalia |
| 17 | 29 | 0.8 | 571   | 21 AAC52152 | Arabidopsis thalia |
| 18 | 29 | 0.8 | 825   | 21 AAC55083 | Arabidopsis thalia |
| 19 | 28 | 0.7 | 621   | 23 AAS90688 | Arabidopsis thalia |
| 20 | 27 | 0.7 | 299   | 23 AAS71084 | DNA encoding novel |
| 21 | 27 | 0.7 | 435   | 23 AAS67160 | DNA encoding novel |
| 22 | 27 | 0.7 | 435   | 23 AAS68115 | DNA encoding novel |
| 23 | 27 | 0.7 | 435   | 23 AAS69544 | DNA encoding novel |
| 24 | 27 | 0.7 | 435   | 23 AAS70705 | DNA encoding novel |
| 25 | 27 | 0.7 | 435   | 23 AAS71149 | DNA encoding novel |
| 26 | 26 | 0.7 | 459   | 21 AAC52482 | Arabidopsis thalia |
| 27 | 26 | 0.7 | 988   | 21 AAC56084 | Eucalyptus grandis |
| 28 | 26 | 0.7 | 1083  | 21 AAC44171 | Arabidopsis thalia |
| 29 | 26 | 0.7 | 1227  | 21 AAC40176 | Arabidopsis thalia |
| 30 | 26 | 0.7 | 86584 | 21 AAF22292 | BAC containing rep |
| 31 | 25 | 0.7 | 140   | 21 AAC13041 | Human secreted pro |
| 32 | 25 | 0.7 | 232   | 21 AAA31626 | Plant microsatelli |
| 33 | 25 | 0.7 | 252   | 18 AAT65064 | Canine genomic mic |
| 34 | 25 | 0.7 | 252   | 20 AAX17801 | Microsatellite rep |
| 35 | 25 | 0.7 | 341   | 21 AAF13992 | Aspergillus oryzae |
| 36 | 25 | 0.7 | 462   | 22 ABA55668 | Human foetal liver |
| 37 | 25 | 0.7 | 462   | 22 ABA35363 | Probe #3829 for ge |
| 38 | 25 | 0.7 | 462   | 22 AAK29365 | Human bone marrow  |
| 39 | 25 | 0.7 | 462   | 22 AAI13948 | Probe #3881 for ge |
| 40 | 25 | 0.7 | 972   | 21 AAC49195 | Arabidopsis thalia |
| 41 | 25 | 0.7 | 1220  | 21 AAC36229 | Arabidopsis thalia |
| 42 | 25 | 0.7 | 1676  | 21 AAC36166 | Arabidopsis thalia |
| 43 | 25 | 0.7 | 2265  | 21 AAC43317 | Arabidopsis thalia |
| 44 | 25 | 0.7 | 2286  | 14 AAQ45913 | Rice starch syntha |
| 45 | 25 | 0.7 | 2340  | 23 AAS86699 | DNA encoding novel |

## ALIGNMENTS

RESULT 1  
AAC74569  
ID AAC74569 standard; cDNA; 333 BP.  
XX  
AC AAC74569;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human ORFX ORF124 polynucleotide sequence SEQ ID NO:247.  
XX  
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnaray; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
KW immunosuppressant; antidiabetic;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW antitumor; dermatological; immunosuppressive; antinflammatory;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antinflammatory disease; coagulation;  
KW thrombosis; contraceptive; ss.  
XX Homo sapiens.  
XX WO200058473-A2.  
XX  
XX 05-OCT-2000.  
XX  
XX 31-MAR-2000; 2000WO-US08621.  
XX PF

[illegible]



```

XX WPI; 1997-202224/18.
XX
XX Ribozyme which modulates plant gene expression - preferably
PT modulates expression of Delta-9 desaturase or granule bound starch
PT synthase in maize or canola
PT
XX Example 9; Page 31-33; 155pp; English.
XX
XX The present invention describes an enzymatic nucleic acid molecule (I)
CC with RNA cleaving activity, which modulates the expression of a plant
CC gene. Also described is a gene comprising a cDNA sequence encoding maize
CC Delta-9 desaturase. (I) can be used to modulate expression of a gene,
CC preferably Delta-9 desaturase or a granule bound starch synthase (GBSS)
CC gene, in a plant (preferably a maize or canola plant). (I) can be used
CC to modulate caffeine synthesis in a coffee plant, nicotine production in
CC a tobacco plant, fruit ripening processes in an apple, tomato, pear,
CC plum or peach plant, flower pigmentation in a rose, petunia,
CC chrysanthemum or marigold plant or lignin production in a tobacco,
CC aspen, poplar or pine plant.
XX
XX Sequence 2267 BP; 465 A; 699 C; 733 G; 370 T; 0 other;
SQ
Query Match 1.08; Score 39; DB 18; Length 2267;
Best Local Similarity 100.0%; Pred. No. 3.3e-08;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3576 cgctacgaccagtagcaagcgcctgggacaccagcgctc 3614
Db 543 cgctacgaccagtagcaagcgcctgggacaccagcgctc 581
RESULT 4
AAV29752
ID AAV29752 standard; DNA; 4800 BP.
XX
XX AAV29752;
XX
XX 11-SEP-1998 (first entry)
XX
XX Zea mays waxy gene.
XX
XX SER: starch-encapsulating region; fusion vector;
KW glucosyl transferase; ss.
XX
XX Zea mays.
XX
XX Key Location/Qualifiers
FH repeat_region /*tag= a
FT /*note= "direct repeat 1"
FT repeat_region 288..292
FT /*tag= b
FT repeat_region 293..297
FT /*tag= c
FT repeat_region 298..302
FT /*tag= d
FT /*note= "direct repeat 1"
FT misc_feature 372..385
FT /*tag= e
FT /*note= "GC stretch, potential regulatory factor binding"
FT misc_feature 442..468
FT /*tag= f
FT /*note= "GC stretch, potential regulatory factor binding"
FT misc_feature 768..782
FT /*tag= g
FT /*note= "GC stretch, potential regulatory factor binding"
FT misc_feature 810..822
FT /*tag= h
FT /*note= "GC stretch, potential regulatory factor binding"
FT

```

---

```

FT misc_feature 821..828
FT /*tag= i
FT /*note= "target duplication site (AC7)"
FT 821..828
FT /*tag= j
FT TATA_signal 867..873
FT /*tag= k
FT misc_feature 887..900
FT /*tag= l
FT /*note= "GC stretch, potential regulatory factor"
FT 901
FT /*tag= m
FT /*note= "transcriptional start site"
FT 1449..14343
FT /*tag= n
FT /*product= glucosyl transferase
FT /*note= "contains introns"
FT 901..1080
FT /*tag= o
FT /*number= 1
FT intron 1081..1219
FT /*tag= p
FT /*number= 1
FT exon 1220..1553
FT /*tag= q
FT /*number= 2
FT intron 1554..1684
FT /*tag= r
FT /*number= 2
FT exon 1685..1765
FT /*tag= s
FT /*number= 3
FT intron 1766..1859
FT /*tag= t
FT /*number= 3
FT exon 1860..1958
FT /*tag= u
FT /*number= 4
FT intron 1959..2054
FT /*tag= v
FT /*number= 4
FT exon 2055..2144
FT /*tag= w
FT /*number= 5
FT intron 2145..2225
FT /*tag= x
FT /*number= 5
FT exon 2226..2289
FT /*tag= y
FT /*number= 6
FT intron 2290..2412
FT /*tag= z
FT /*number= 6
FT exon 2413..2513
FT /*tag= aa
FT /*number= 7
FT intron 2514..2650
FT /*tag= bb
FT /*number= 7
FT exon 2651..2760
FT /*tag= cc
FT /*number= 8
FT intron 2761..2857
FT /*tag= dd
FT /*number= 8
FT exon 2858..3101
FT /*tag= ee
FT /*number= 9
FT intron 3102..3211
FT /*tag= ff
FT /*number= 9
FT exon 3212..3394
FT /*tag= gg
FT

```

FT misc\_feature /number= 10  
FT 3358..3489 /tag= hh  
FT /note= "target duplication site (Ac9)"  
FT intron /tag= 11  
FT 3395..3489 /number= 10  
FT /tag= 10  
FT exon 3490..3681 /tag= jj  
FT /number= 11  
FT misc\_feature /number= 11  
FT 3570..3572 /tag= kk  
FT /note= "target duplication site (Spm 18)"  
FT intron 3682..3792 /tag= ll  
FT /number= 11  
FT exon 3793..3879 /tag= mm  
FT /number= 12  
FT intron 3880..3976 /tag= nn  
FT /number= 12  
FT exon 3977..4105 /tag= oo  
FT /number= 13  
FT intron 4106..4226 /tag= pp  
FT /number= 13  
FT exon 4227..4595 /tag= qq  
FT /number= 14  
FT polyA\_signal 4570..4575 /tag= rr  
FT polyA\_signal 4593..4598 /tag= ss  
FT polyA\_signal 4597..4602 /tag= tt  
FT WO9814601-A1.  
XX  
XX  
XX 09-APR-1998.  
XX 30-SEP-1997; 97WO-US17555.  
XX 30-SEP-1996; 96US-0026855.  
XX (EXSE-) EXSEED GENETICS LLC.  
XX Guan H, Keeling P;  
XX WPI: 1998-240100/21.  
XX P-PSDB; AAW56484.  
XX Hybrid polypeptide comprising starch-encapsulating region and  
XX protein - useful for, e.g. producing protein(s) resistant to  
XX degradation by stomach acids  
XX Example 2; Page 29-31; 156pp; English.  
XX The sequence is that of the waxy gene which codes for glucosyl  
XX transferase. It can be used in the production of a hybrid  
XX polypeptide comprising a starch-encapsulating region (SER) fused  
XX to a payload protein. The hybrid polypeptide can be used to make  
XX modified starches comprising the payload protein, selected from,  
XX e.g. hormones, growth factors, antibodies, enzymes, dyes,  
XX immunoglobulins, etc. The modified starch can also be used  
XX to provide grain feeds enriched in amino acids. By encapsulating  
XX the payload protein in starch, it is more resistant to  
XX degradation by stomach acids.  
XX Sequence 4800 BP; 935 A; 1414 C; 1446 G; 1005 T; 0 other;

Query Match 1.0%; Score 39; DB 19; Length 4800;  
Best Local Similarity 100.0%; Pred. No. 3.3e-08;  
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 3576 cgctagaccagctacagcagcctgggacacacgctc 3614  
Db 1718 cgctagaccagctacagcagcctgggacacacgctc 1756  
RESULT 5  
ABA68266/C  
ID ABA68266 standard; DNA; 276 BP.  
XX ABA68266;  
AC ABA68266;  
XX 01-FEB-2002 (first entry)  
XX Human foetal liver single exon nucleic acid probe #16571.  
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
XX Homo sapiens.  
XX WO200157277-A2.  
XX 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US00669.  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI: 2001-483447/52.  
XX Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human fetal liver -  
XX Claim 4; SEQ ID NO 16571; 639pp + sequence listing; English.  
XX The invention relates to a single exon nucleic acid probe for  
XX measuring human gene expression in a sample derived from human foetal  
XX liver. The single exon nucleic acid probes may be used for predicting,  
XX measuring and displaying gene expression in samples derived from human  
XX foetal liver. The present sequence is a single exon nucleic acid  
XX probe of the invention.  
XX Note: The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 276 BP; 23 A; 81 C; 20 G; 152 T; 0 other;  
SQ  
Query Match 1.0%; Score 37; DB 22; Length 276;  
Best Local Similarity 100.0%; Pred. No. 3e-07;  
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 2379 gaagagaagaagaagaagaagaagaagaagaagaag 2415  
Db 243 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 207  
RESULT 6  
ABA35275/C  
ID ABA35275 standard; DNA; 276 BP.









CC present in interval 6D and/or 6E of the distal portion of Yq, mutations  
CC in which are associated with reduced sperm count. The DAZH gene  
CC (AAZ92580) is located on chromosome 3; however, the entire DAZ gene  
CC family, including DAZH is expressed in germ cells. DAZ and DAZH  
CC nucleotide sequences may be used as a source of primers and probes for  
CC the diagnosis of cases of reduced sperm count associated with alteration  
CC or deletion of the DAZ gene. They are also used as human chromosome Y  
CC markers. Functional DAZ genes can be used in gene therapy for treating  
CC reduced sperm counts. Sequence AAZ92499 represents human DAZ cDNA, and  
CC sequences AAZ92583-292584 are genomic DAZ sequences. Sequences  
CC AAZ92491-292492 are partial DNA sequences of DAZ gene family members.  
XX  
SQ Sequence 43795 BP; 12175 A; 8166 C; 8183 G; 15271 T; 0 other;

Query Match 0.8%; Score 31; DB 21; Length 43795;  
Best Local Similarity 100.0%; Pred. No. 0.00023;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2385 gaagaagaagaagcagagaagaagaag 2415  
|||||  
Db 38989 gaagaagaagaagcagagaagaagaag 39019

## RESULT 15

AAC49802  
ID AAC49802 standard; DNA; 509 BP.

XX AC AAC49802;

XX XX 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 62490.

KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.

XX OS Arabidopsis thaliana.

PN EP1033405-A2.

XX PD 06-SEP-2000.

XX XX 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145195.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 0.8%; Score 29; DB 21; Length 509;

Best Local Similarity 100.0%; Pred. No. 0.0021;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 aagaagaagaagaagaagaagaagaagaaga 2411

||||||||||||||||||||||||||||||||

Db 115 aagaagaagaagaagaagaagaagaagaaga 143

Search completed: July 31, 2002, 16:20:22

Job time: 8259 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:26:27 ; Search time 165.21 Seconds  
(without alignments)  
26.762 Million cell updates/sec

Title: US-09-899-718A-8  
Perfect score: 18  
Sequence: 1 atactctggtaagtttaa 18

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, NA: \*  
1: /cgn2.6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2.6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2.6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2.6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2.6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2.6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| C 1        | 14.4  | 80.0        | 890    | 3  | US-08-741-411-4    |
| 2          | 14.4  | 80.0        | 1070   | 1  | Sequence 4, Appli  |
| 3          | 14.4  | 80.0        | 1070   | 4  | Sequence 11, Appl  |
| 4          | 14.4  | 80.0        | 1410   | 3  | Sequence 11, Appl  |
| 5          | 14.4  | 80.0        | 1410   | 4  | Sequence 1, Appli  |
| 6          | 14.4  | 80.0        | 3533   | 2  | Sequence 1, Appli  |
| 7          | 14.4  | 80.0        | 3533   | 5  | Sequence 40, Appl  |
| 8          | 14.4  | 80.0        | 3533   | 6  | Sequence 40, Appl  |
| 9          | 13.8  | 76.7        | 248    | 4  | Patent No. 5424399 |
| 10         | 13.8  | 76.7        | 248    | 4  | Sequence 204, App  |
| 11         | 13.8  | 76.7        | 908    | 4  | Sequence 204, App  |
| 12         | 13.8  | 76.7        | 3147   | 4  | Sequence 15, Appl  |
| 13         | 13.8  | 76.7        | 3180   | 1  | Sequence 12, Appl  |
| 14         | 13.8  | 76.7        | 3180   | 3  | Sequence 1, Appli  |
| 15         | 13.8  | 76.7        | 3180   | 4  | Sequence 1, Appli  |
| 16         | 13.8  | 76.7        | 3180   | 5  | Sequence 1, Appli  |
| 17         | 13.8  | 76.7        | 3930   | 2  | Sequence 1, Appli  |
| 18         | 13.8  | 76.7        | 12730  | 4  | Sequence 2, Appli  |
| 19         | 13.8  | 76.7        | 15062  | 4  | Sequence 91, Appl  |
| 20         | 13.4  | 74.4        | 17     | 1  | Sequence 89, Appl  |
| 21         | 13.4  | 74.4        | 17     | 1  | Sequence 1713, Ap  |
| 22         | 13.4  | 74.4        | 825    | 1  | Sequence 1713, Ap  |
| 23         | 13.4  | 74.4        | 825    | 4  | Sequence 48, Appl  |
| 24         | 13.4  | 74.4        | 1234   | 1  | Sequence 48, Appl  |
| 25         | 13.4  | 74.4        | 1328   | 4  | Sequence 1, Appli  |
| 26         | 13.4  | 74.4        | 6028   | 4  | Sequence 71, Appl  |
| 27         | 13.4  | 74.4        | 8202   | 1  | Sequence 5, Appli  |
|            |       |             |        |    | Sequence 13, Appl  |

|      |      |      |       |   |                   |                    |
|------|------|------|-------|---|-------------------|--------------------|
| 28   | 13.4 | 74.4 | 14176 | 1 | US-08-307-499-1   | Sequence 1, Appli  |
| C 29 | 13.4 | 74.4 | 14176 | 1 | US-08-307-499-14  | Sequence 14, Appl  |
| 30   | 13.4 | 74.4 | 14176 | 4 | US-09-299-268-1   | Sequence 1, Appli  |
| C 31 | 13.4 | 74.4 | 14176 | 4 | US-09-299-268-14  | Sequence 14, Appl  |
| C 32 | 13.2 | 73.3 | 24    | 2 | US-08-680-327-6   | Sequence 6, Appli  |
| 33   | 13.2 | 73.3 | 132   | 1 | US-08-248-474-5   | Sequence 5, Appli  |
| 34   | 13.2 | 73.3 | 132   | 3 | US-08-756-849-5   | Sequence 5, Appli  |
| C 35 | 13.2 | 73.3 | 455   | 6 | 5278286-1         | Patent No. 5278286 |
| C 36 | 13.2 | 73.3 | 601   | 2 | US-08-184-009-168 | Sequence 168, App  |
| C 37 | 13.2 | 73.3 | 601   | 2 | US-08-458-356-168 | Sequence 168, App  |
| C 38 | 13.2 | 73.3 | 601   | 4 | US-08-460-736-168 | Sequence 168, App  |
| 39   | 13.2 | 73.3 | 736   | 3 | US-08-771-110-1   | Sequence 1, Appli  |
| C 40 | 13.2 | 73.3 | 1095  | 2 | US-08-625-377-3   | Sequence 3, Appli  |
| C 41 | 13.2 | 73.3 | 1095  | 4 | US-09-204-841-3   | Sequence 3, Appli  |
| C 42 | 13.2 | 73.3 | 1251  | 4 | US-09-242-095-1   | Sequence 1, Appli  |
| 43   | 13.2 | 73.3 | 1452  | 1 | US-08-149-093A-3  | Sequence 3, Appli  |
| 44   | 13.2 | 73.3 | 1452  | 1 | US-08-911-245-3   | Sequence 3, Appli  |
| 45   | 13.2 | 73.3 | 1452  | 1 | US-08-553-058C-3  | Sequence 3, Appli  |

## ALIGNMENTS

RESULT 1  
US-08-741-411-4/c  
; Sequence 4, Application US/08741411  
; Patent No. 6124116  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; TITLE OF INVENTION: NOVEL RAB PROTEINS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/741.411  
; FILING DATE: Herewith  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0139 US  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 890 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; IMMEDIATE SOURCE:  
; LIBRARY: Consensus  
; CLONE: Consensus  
US-08-741-411-4

Query Match 80.0%; Score 14.4; DB 3; Length 890;

Best Local Similarity 93.8%; Pred. No. 47;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctgttcattt 16  
||||| |||||||  
Db 282 ATACTCGGTCATGTT 267

## RESULT 2

US-08-602-713-11  
; Sequence 11, Application US/08602713  
; Patent No. 5798205  
; GENERAL INFORMATION:  
; APPLICANT: Hauser, Hans-Peter; Knapp Stefan; Brust, Stefan;  
; APPLICANT: G rler, Lutz G.; Eberle, Josef; Kaptue, Lazare;  
; APPLICANT: Zekeng, L Opold Achengu  
; TITLE OF INVENTION: Retrovirus From The HIV Group And Its Use  
; TITLE OF INVENTION: (MVP-2901/94)  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Felie & Lynch  
; STREET: 805 Third Avenue  
; CITY: New York City  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/602.713  
; FILING DATE: 16-FEBRUARY-1996  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; FILING DATE: 16-FEBRUARY-1995  
; APPLICATION NUMBER: DE 195 05 262  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hanson, No. 5798205man D.  
; REGISTRATION NUMBER: 30,946  
; REFERENCE/DOCKET NUMBER: LEDE 203  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 688-9200  
; TELEFAX: (212) 838-3884  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1070 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
US-08-602-713-11

Query Match 80.0%; Score 14.4; DB 1; Length 1070;  
Best Local Similarity 93.8%; Pred. No. 48;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctgttcattt 16  
||||| |||||||  
Db 403 ATACTCTGTTCATGTT 418

## RESULT 3

US-08-989-493-11  
; Sequence 11, Application US/08989493  
; Patent No. 6162631  
; GENERAL INFORMATION:  
; APPLICANT: Hauser, Hans-Peter; Knapp Stefan; Brust, Stefan;  
; APPLICANT: G rler, Lutz G.; Eberle, Josef; Kaptue, Lazare;  
; APPLICANT: Zekeng, L Opold Achengu  
; TITLE OF INVENTION: Retrovirus From The HIV Group And Its Use

; TITLE OF INVENTION: (MVP-2901/94)  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Felie & Lynch  
; STREET: 805 Third Avenue  
; CITY: New York City  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/989.493  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/602.713  
; FILING DATE: 16-FEBRUARY-1996  
; APPLICATION NUMBER: DE 195 05 262  
; FILING DATE: 16-FEBRUARY-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hanson, No. 6162631man D.  
; REGISTRATION NUMBER: 30,946  
; REFERENCE/DOCKET NUMBER: LEDE 203  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 688-9200  
; TELEFAX: (212) 838-3884  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1070 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
US-08-989-493-11

Query Match 80.0%; Score 14.4; DB 4; Length 1070;  
Best Local Similarity 93.8%; Pred. No. 48;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctgttcattt 16  
||||| |||||||  
Db 403 ATACTCTGTTCATGTT 418

## RESULT 4

US-08-147-592A-1  
; Sequence 1, Application US/08147592A  
; Patent No. 6096513  
; GENERAL INFORMATION:  
; APPLICANT: Bell, Graeme I  
; APPLICANT: Reisine, Terry  
; APPLICANT: Yasuda, Kazuki  
; TITLE OF INVENTION: Opioid Receptor Genes,  
; TITLE OF INVENTION: Compositions and Methods  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: United States of America  
; ZIP: 77210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

Sequence 40, Application US/084/6062A  
Patent No. 5877275  
GENERAL INFORMATION:  
APPLICANT: Arnaout, M. Amin  
TITLE OF INVENTION: CONTROLLING CELLULAR IMMUNE/INFLAMMATORY  
TITLE OF INVENTION: RESPONSES WITH BETA2 INTEGRINS  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/476,062A  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/216,081  
FILING DATE: 21-MAR-1994  
APPLICATION NUMBER: 07/637,830  
FILING DATE: 04-JAN-1991  
APPLICATION NUMBER: 07/539,842  
FILING DATE: 18-JUN-1990  
APPLICATION NUMBER: 07/212,573  
FILING DATE: 28-JUN-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Freeman, John W.  
REGISTRATION NUMBER: 29, 066  
REFERENCE/DOCKET NUMBER: 00786/068003

Query Match 80.0%; Score 14.4; DB 5; Length 3533;  
Best Local Similarity 93.8%; Pred. No. 56;  
Matches 15; Conservative 0; Mismatches 1; Indels 0

INFORMATION FOR SEQ ID NO: 204:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 248 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

us-09-899-718a-8.rni

Thu Aug 1 08:30:07 2002

```

; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 204:
US-08-991-789A-204

Query Match          76.7%; Score 13.8; DB 4; Length 248;
Best Local Similarity 88.2%; Pred. No. 83;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atactctggtcatgtta 17
   || ||||| ||||| |||||
Db 71 ATGCTCTGATCATGTGA 87

RESULT 10
US-09-062-451-204
; Sequence 204, Application US/09062451
; Patent No. 6344550
; GENERAL INFORMATION:
; APPLICANT: Fridakis, Tony N.
; APPLICANT: Smith, John M.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 297
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/062,451
; FILING DATE: 04-APR-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.419C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 204:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 248 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-062-451-204

Query Match          76.7%; Score 13.8; DB 4; Length 248;
Best Local Similarity 88.2%; Pred. No. 83;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atactctggtcatgtta 17
   || ||||| ||||| |||||
Db 71 ATGCTCTGATCATGTGA 87

RESULT 11
US-09-457-046B-15
; Sequence 15, Application US/09457046B
; Patent No. 6287835
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway
; FILE REFERENCE: 53679
; CURRENT APPLICATION NUMBER: US/09/457,046B
; CURRENT FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 15
; LENGTH: 908
; TYPE: DNA
; ORGANISM: Taxus cuspidata
US-09-457-046B-15

Query Match          76.7%; Score 13.8; DB 4; Length 908;
Best Local Similarity 88.2%; Pred. No. 99;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atactctggtcatgtta 17
   ||||| ||||| |||||
Db 663 atactatggttaattgta 679

RESULT 12
US-09-153-804-12
; Sequence 12, Application US/09153804
; Patent No. 6207380
; GENERAL INFORMATION:
; APPLICANT: Patricia Billing-Medel
; APPLICANT: Maurice Cohen
; APPLICANT: Paula N. Friedman
; APPLICANT: Julian Gordon
; APPLICANT: Steven C. Hodges
; APPLICANT: Michael R. Klass
; APPLICANT: Jon D. Kratochvil
; APPLICANT: Eric Russell
; APPLICANT: Stephen D. Stroupe
; TITLE OF INVENTION: Reagents and Methods Useful for Detecting Diseases of the Urin
; FILE REFERENCE: 6180.US.01
; CURRENT APPLICATION NUMBER: US/09/153,804
; CURRENT FILING DATE: 1998-09-15
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-153-804-12

Query Match          76.7%; Score 13.8; DB 4; Length 3147;
Best Local Similarity 88.2%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tactctggtcatgttaa 18
   | ||||| ||||| |||||
Db 1336 ttctctggttattgtaa 1352

RESULT 13
US-08-480-662-1
; Sequence 1, Application US/08480662
; Patent No. 5759782
; GENERAL INFORMATION:
; APPLICANT: Pastan, Ira
; APPLICANT: Brinkmann, Ulrich
; TITLE OF INVENTION: CELLULAR APOPTOSIS SUSCEPTIBILITY PROTEIN (CSP) AND AN
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson and Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA

```

ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,662  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned A  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH112.001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-235-8550  
TELEFAX: 619-235-0176  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3180 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
US-08-480-662-1

Query Match 76.7%; Score 13.8; DB 1; Length 3180;  
Best Local Similarity 88.2%; Pred. No. 1.2e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tactctggatcatgtaa 18  
Db 1336 TTCTCTGGTATGTTAA 1352

RESULT 14  
US-08-918-190-1  
Sequence 1, Application US/08918190  
Patent No. 6072031  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira  
TITLE OF INVENTION: CELLULAR APOPTOSIS SUSCEPTIBILITY  
TITLE OF INVENTION: PROTEIN (CSP) AND ANTISENSE CSP  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson and Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: USA  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/918,190  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/480,662  
FILING DATE: 07-JUN-1995

ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned A  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH112.001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-235-8550  
TELEFAX: 619-235-0176  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3180 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
US-08-918-190-1

Query Match 76.7%; Score 13.8; DB 3; Length 3180;  
Best Local Similarity 88.2%; Pred. No. 1.2e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tactctggatcatgtaa 18  
Db 1336 TTCTCTGGTATGTTAA 1352

RESULT 15  
US-09-234-232-1  
Sequence 1, Application US/09234232  
Patent No. 6232086  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira  
Brinkmann, Ulrich  
TITLE OF INVENTION: CELLULAR APOPTOSIS SUSCEPTIBILITY  
PROTEIN (CSP) AND ANTISENSE CSP  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson and Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: USA  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/234,232  
FILING DATE: 20-Jan-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/918,190  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned A  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH112.001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-235-8550  
TELEFAX: 619-235-0176  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3180 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single



us-09-899-718a-8.rni

Thu Aug 1 08:30:07 2002

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-234-232-1

Query Match      76.7%; Score 13.8; DB 4; Length 3180;
Best Local Similarity 88.2%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 tactctggtcatgttaa 18
        | ||||| |||||
Db 1336 TTCCTGCTTATGTAA 1352

Search completed: July 31, 2002, 12:26:28
Job time: 12011 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:12:12 ; Search time 5855.71 Seconds  
(without alignments)  
41.489 Million cell updates/sec

Title: US-09-899-718a-8  
Perfect score: 18  
Sequence: 1 atactctgtcatgttaa 18

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID        | Description         |
|------------|-------|-------------|--------|----|-----------|---------------------|
| C 1        | 17    | 94.4        | 1441   | 10 | BG531376  | BG531376 602559625  |
| C 2        | 16.4  | 91.1        | 344    | 10 | BG516695  | BG516695 NXSL 002   |
| C 3        | 16.4  | 91.1        | 513    | 10 | BG318395  | BG318395 NXPV 013   |
| C 4        | 16.4  | 91.1        | 558    | 12 | AQ621441  | AQ621441 HS_3019_B  |
| C 5        | 16.4  | 91.1        | 564    | 12 | AQ621441  | AQ621441 HS_3019_B  |
| C 6        | 16.4  | 91.1        | 571    | 10 | BG318023  | BG318023 NXPV 008   |
| C 7        | 16.4  | 91.1        | 724    | 12 | AZ207084  | AZ207084 SP_0127_A  |
| C 8        | 16    | 88.9        | 404    | 12 | AQ150695  | AQ150695 HS_3203_A  |
| C 9        | 16    | 88.9        | 635    | 9  | AW812504  | AW812504 CM4-ST018  |
| C 10       | 15.6  | 86.7        | 1026   | 12 | CNS0275Q  | AL212759 Tetraodon  |
| C 11       | 15.4  | 85.6        | 213    | 10 | BE1287040 | BE1287040 UI-R-Cf05 |
| C 12       | 15.4  | 85.6        | 225    | 9  | BE142340  | BE142340 CM3-HT014  |
| C 13       | 15.4  | 85.6        | 297    | 9  | BB024447  | BB024447 BB024447   |
| C 14       | 15.4  | 85.6        | 303    | 9  | BB289468  | BB289468 BB289468   |
| C 15       | 15.4  | 85.6        | 337    | 10 | BF147142  | BF147142 uy07f02.Y  |
| C 16       | 15.4  | 85.6        | 350    | 10 | BE556305  | BE556305 sq01a05.Y  |
| C 17       | 15.4  | 85.6        | 393    | 10 | BE650684  | BE650684 UI-M-BH3-  |

|      |      |      |     |    |          |                    |
|------|------|------|-----|----|----------|--------------------|
| C 18 | 15.4 | 85.6 | 410 | 12 | AQ442997 | AQ442997 HS_5132_B |
| C 19 | 15.4 | 85.6 | 425 | 9  | AA102952 | AA102952 mo09d01.r |
| C 20 | 15.4 | 85.6 | 443 | 10 | BF256005 | BF256005 HVSMEF000 |
| C 21 | 15.4 | 85.6 | 451 | 10 | BI326958 | BI326958 AR071D101 |
| C 22 | 15.4 | 85.6 | 457 | 9  | AW738969 | AW738969 gb16h09.Y |
| C 23 | 15.4 | 85.6 | 476 | 9  | AA152987 | AA152987 mq48c03.r |
| C 24 | 15.4 | 85.6 | 481 | 9  | AI493734 | AI493734 qz12g07.x |
| C 25 | 15.4 | 85.6 | 494 | 9  | AW230662 | AW230662 uo6d09.Y  |
| C 26 | 15.4 | 85.6 | 495 | 12 | AQ427545 | AQ427545 CITBI-E1- |
| C 27 | 15.4 | 85.6 | 505 | 10 | HI6901   | HI6901 ym39b02.r1  |
| C 28 | 15.4 | 85.6 | 512 | 9  | AW395872 | AW395872 sh01c11.Y |
| C 29 | 15.4 | 85.6 | 529 | 10 | C84686   | C84686 C84686 Dict |
| C 30 | 15.4 | 85.6 | 537 | 12 | AQ596806 | AQ596806 HS_5206_A |
| C 31 | 15.4 | 85.6 | 543 | 10 | BI784821 | BI784821 saf93g09. |
| C 32 | 15.4 | 85.6 | 552 | 12 | A246415  | A246415 IM0242014  |
| C 33 | 15.4 | 85.6 | 558 | 10 | BI784791 | BI784791 saf93g11. |
| C 34 | 15.4 | 85.6 | 564 | 9  | AW232748 | AW232748 f129g04.x |
| C 35 | 15.4 | 85.6 | 571 | 12 | AQ802955 | AQ802955 HS_3160_A |
| C 36 | 15.4 | 85.6 | 617 | 12 | AQ787071 | AQ787071 HS_3156_B |
| C 37 | 15.4 | 85.6 | 629 | 12 | AG017311 | AG017311 Homo sapi |
| C 38 | 15.4 | 85.6 | 634 | 12 | AG058023 | AG058023 Pan trogl |
| C 39 | 15.4 | 85.6 | 635 | 10 | BE369456 | BE369456 601220868 |
| C 40 | 15.4 | 85.6 | 635 | 12 | AQ018921 | AQ018921 CIT-HSP-2 |
| C 41 | 15.4 | 85.6 | 637 | 12 | AG017310 | AG017310 Homo sapi |
| C 42 | 15.4 | 85.6 | 646 | 12 | AG017312 | AG017312 Homo sapi |
| C 43 | 15.4 | 85.6 | 673 | 12 | BH094331 | BH094331 RPCI-24-3 |
| C 44 | 15.4 | 85.6 | 688 | 10 | BG694687 | BG694687 NISC_1v06 |
| C 45 | 15.4 | 85.6 | 697 | 9  | AV836265 | AV836265 AV836265  |

## ALIGNMENTS

RESULT 1  
BG531376/c

LOCUS 602559625F1 NIH\_MGC\_61 Homo sapiens cDNA clone IMAGE:4697383 5',  
DEFINITION mRNA sequence.  
ACCESSION BG531376  
VERSION BG531376.1 GI:13522913  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1441)  
AUTHORS NTH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC

CDNA Library Preparation: CLONETECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM1526 row: b column: 08  
High quality sequence start: 186  
High quality sequence stop: 226.  
Location/Qualifiers  
1. 1441  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4697383"  
/clone.lib="NIH\_MGC\_61"  
/tissue\_type="embryonal carcinoma"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site:1:  
SfiI (ggccgcctcgcc); Site:2: SfiI (ggccattatggcc);  
Double-stranded cDNA was prepared from cell line RNA.  
and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CACGCCATTATGCCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGCGCCGACATG-DT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC Library."

BASE COUNT 413 a 381 c 367 g 279 t 1 others  
ORIGIN

Query Match 94.4%; Score 17; DB 10; Length 1441;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atactctggatcgttta 17  
|||||

Db 183 ATACTCTGTCATGTTA 167

## RESULT 2

BF516695 344 bp mRNA linear EST 08-DEC-2000  
LOCUS NXSI\_002\_C09\_F NXSI (Nsf Xylem side wood Inclined) Pinus taeda cDNA  
DEFINITION clone NXSI\_002\_C09 5', mRNA sequence.  
ACCESSION BF516695  
VERSION BF516695.1 GI:11603809  
KEYWORDS EST.  
SOURCE loblolly pine.  
ORGANISM Pinus taeda

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Coniferales; Pinaceae; Pinus; Pinus.  
1 (bases 1 to 344)  
Sederoff, R.

TITLE Molecular Basis of Wood Formation in the Pine Megagenome  
JOURNAL Unpublished (2000)  
COMMENT Contact: Johnson, Arthur  
North Carolina State University

Tel: 919 515 7800  
Fax: 919 515 7801  
Email: ajohnson@unity.ncsu.edu  
Seq primer: T3.

## FEATURES

source  
Location/Qualifiers  
1..344  
/organism="Pinus taeda"  
/strain="Coastal plain loblolly pine from North Carolina"  
/db\_xref="taxon:3352"  
/clone="NXSI\_002\_C09"  
/tissue\_type="xylem"  
/cell\_type="side"  
/dev\_stage="juvenile"  
/lab\_host="X11-Blue"  
/note="Vector: Bluescript SK; Site\_1: Eco RI; Site\_2: XhoI  
; The library is from early (spring) wood, taken from  
three six-year old trees (three different genotypes), in  
the juvenile phase. These trees were induced to form side  
wood by bending to a 45 degree angle and tying them to the  
ground. Differentiating xylem was harvested from the sides  
of the inclined stems, and a mixture of all three  
genotypes was used for the library. oligo-dr primed cDNA  
was directionally cloned into the EcoRI-XhoI Bluescript SK  
vector arms. NOTE: The sequences contain a 'cDNA adapter'  
between the EcoRI site and the start of the EST. The  
adapter sequence is 'AATTCGGCAGCAG'."

BASE COUNT 80 a 74 c 75 g 103 t 12 others  
ORIGIN

Query Match 91.1%; Score 16.4; DB 10; Length 344;  
Best Local Similarity 94.4%; Pred. No. 4.1e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggatcgttta 18  
|||||

Db 237 ATACTCTGTCATGTTAA 254

## RESULT 3

BF318395 513 bp mRNA linear EST 26-FEB-2001  
LOCUS NXPV\_013\_B06\_F NXPV (Nsf Xylem Planings wood Vertical) Pinus taeda  
DEFINITION cDNA clone NXPV\_013\_B06 5', mRNA sequence.

ACCESSION BF318395  
VERSION BF318395.1 GI:13127825  
KEYWORDS EST.  
SOURCE loblolly pine.  
ORGANISM Pinus taeda

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Coniferales; Pinaceae; Pinus; Pinus.  
1 (bases 1 to 513)  
Sederoff, R.

TITLE Molecular Basis of Wood Formation in the Pine Megagenome  
JOURNAL Unpublished (2000)  
COMMENT Contact: Johnson, Arthur  
North Carolina State University

Tel: 919 515 7800  
Fax: 919 515 7801  
Email: ajohnson@unity.ncsu.edu  
Seq primer: T3.

## FEATURES

source  
Location/Qualifiers  
1..513  
/organism="Pinus taeda"  
/strain="Coastal plain loblolly pine from North Carolina"  
/db\_xref="taxon:3352"  
/clone="NXPV\_013\_B06"  
/tissue\_type="xylem"  
/cell\_type="planings (secondary)"  
/dev\_stage="transitional"  
/lab\_host="X11-Blue"  
/note="Vector: Bluescript SK; Site\_1: Eco RI; Site\_2: XhoI  
; The library is from early (spring) secondary wood, taken  
from a ten year old tree in the transitional phase. The  
tree is a kind gift of the Westvaco Corporation. Secondary  
xylem was harvested from the tree by peeling back the bark  
and primary xylem and then removing the underlying tissue  
with a block plane. NOTE: The sequences contain a 'cDNA  
adapter' between the EcoRI site and the start of the EST.  
The adapter sequence is 'AATTCGGCAGCAG'."

BASE COUNT 125 a 102 c 117 g 150 t 19 others  
ORIGIN

Query Match 91.1%; Score 16.4; DB 10; Length 513;  
Best Local Similarity 94.4%; Pred. No. 4.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggatcgttta 18  
|||||

Db 324 ATACTCTGTCATGTTAA 341

## RESULT 4

AQ621441/c 558 bp DNA linear GSS 16-JUN-1999  
LOCUS HS\_3019\_B2\_F05\_T7 C17 Approved Human Genomic Sperm Library D Homo  
DEFINITION sapiens genomic clone Plate-3019 Col=10 Row=L, DNA sequence.

ACCESSION AQ621441  
VERSION AQ621441.1 GI:5083833  
KEYWORDS GSS.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

REFERENCE
AUTHORS      1 (bases 1 to 558)
              Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
              Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
              Hood,L.
TITLE        Sequence-tagged connectors: A sequence approach to mapping and
              scanning the human genome
JOURNAL      Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE      99380589
COMMENT      Contact: Mahairas GG, Wallace JC, Hood L
              High Throughput Sequencing Center
              University of Washington
              401 Queen Anne Avenue North, Seattle, WA 98109, USA
              Tel: (206) 616-3618
              Fax: (206) 616-3887
              Email: jwallace@u.washington.edu
              Clones may be purchased from Research Genetics (info@resgen.com).
              BAC end Web Server: http://www.htsc.washington.edu
              Plate: 3019 row: L column: 10
              Seq primer: '7'
              Class: BAC ends
              High quality sequence stop: 558.
              Location/Qualifiers
                1..558
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="Plate-3019 Col-10 Row=L"
                  /clone_lib="CIR Approved Human Genomic Sperm Library D"
                  /sex="male"
                  /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                    E-Coli DH10B"
BASE COUNT   145 a 142 c 118 g 150 t 3 others
ORIGIN
Query Match 91.1%; Score 16.4; DB 12; Length 558;
Best Local Similarity 94.4%; Pred. No. 4.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atactctggtcattgttaa 18
    ||||| |||||
Db 437 ATACTCTGCTCATGTAA 420

RESULT 5
AZ144687 564 bp DNA linear GSS 28-AUG-2000
LOCUS    SP_0041_B2_F12.SP6E Strongylocentrotus purpuratus, purple sea
DEFINITION urchin, sperm genomic BAC library Strongylocentrotus purpuratus
            genomic clone Plate=41 Col=24 Row=L, DNA sequence.
ACCESSION AZ144687
VERSION   AZ144687.1 GI:8296590
KEYWORDS  GSS.
SOURCE    Strongylocentrotus purpuratus.
ORGANISM  Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinozoa; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
            1 (bases 1 to 564)
            Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
            ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and
            Hood,L.
            A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
JOURNAL    Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
MEDLINE    20402566
COMMENT    Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu

```

```

Plate: 41 row: L column: 24
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 564.
Location/Qualifiers
  1..564
    /organism="Strongylocentrotus purpuratus"
    /db_xref="taxon:7668"
    /clone="Plate=41 Col=24 Row=L"
    /clone_lib="Strongylocentrotus purpuratus, purple sea
    urchin, sperm genomic BAC library"
    /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
    DH10B"
BASE COUNT   188 a 97 c 127 g 146 t 6 others
ORIGIN
Query Match 91.1%; Score 16.4; DB 12; Length 564;
Best Local Similarity 94.4%; Pred. No. 4.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atactctggtcattgttaa 18
    ||||| |||||
Db 253 ATACTTGGTCATGTAA 270

RESULT 6
BG318023 571 bp mRNA linear EST 26-FEB-2001
LOCUS    NXPV_008_F06_F NXPV (Nsf Xylem Planings wood Vertical) Pinus taeda
DEFINITION cDNA clone NXPV_008_F06 5', mRNA sequence.
ACCESSION BG318023
VERSION   BG318023.1 GI:13127453
KEYWORDS  EST.
SOURCE    lobliolly pine.
ORGANISM  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
            1 (bases 1 to 571)
            Sederoff,R.
            Molecular Basis of Wood Formation in the Pine Megagenome
            Unpublished (2000)
            Contact: Johnson, Arthur
            North Carolina State University
            Tel: 919 515 7800
            Fax: 919 515 7801
            Email: ajohnson@unity.ncsu.edu
            Seq primer: T3.
            Location/Qualifiers
              1..571
                /organism="Pinus taeda"
                /strain="Coastal plain lobliolly pine from North Carolina"
                /db_xref="taxon:3352"
                /clone="NXPV_008_F06"
                /clone_lib="NXPV (Nsf Xylem Planings wood Vertical)"
                /tissue_type="Xylem"
                /cell_type="Planings (secondary)"
                /dev_stage="Transitional"
                /lab_host="XLI-Blue"
                /note="Vector: Bluescript SK; Site_1: Eco RI; Site_2: XhoI
                ; The library is from early (Spring) secondary wood, taken
                from a ten year old tree in the transitional phase. The
                tree is a kind gift of the Westvaco Corporation. Secondary
                xylem was harvested from the tree by peeling back the bark
                and primary xylem and then removing the underlying tissue
                with a block plane. NOTE: The sequences contain a 'cDNA
                adapter' between the EcoRI site and the start of the EST.
                The adapter sequence is 'AATTGGCAGAG'."
BASE COUNT   141 a 111 c 129 g 169 t 21 others
ORIGIN
Query Match 91.1%; Score 16.4; DB 10; Length 571;

```

```

Best Local Similarity 94.4%; Pred. No. 4.7e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atactctggtcatgttaa 18
    ||||| ||||| ||||| |||||
Db 324 ATACACTGGTCATGTAA 341

RESULT 7
AZ207084/c
LOCUS
DEFINITION
SP_0127_A2_G12_T7A Strongylocentrotus purpuratus, purple sea urchin
    , sperm genomic BAC library Strongylocentrotus purpuratus genomic
    clone Plate-127 Col-24 Row=M, DNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Strongylocentrotus purpuratus.
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinoidea; Echinacea; Echinoidea;
Strongylocentrotidae; Strongylocentrotus.
REFERENCE
1 (bases 1 to 724)
AUTHORS
Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and
Hood,L.
A sea urchin genome project: Sequence scan, virtual map, and
additional resources
Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
20402566
Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 127 Row: M Column: 24
Seq primer: T7
Class: BAC ends
High quality sequence stop: 724.
Location/Qualifiers
1. .724
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone_lib="Strongylocentrotus purpuratus, purple sea
    urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
    DH10B"
BASE COUNT
198 a 149 c 103 g 274 t
ORIGIN

FEATURES
source
1. .724
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone_lib="Strongylocentrotus purpuratus, purple sea
    urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
    DH10B"
BASE COUNT
198 a 149 c 103 g 274 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 12; Length 724;
Best Local Similarity 94.4%; Pred. No. 5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atactctggtcatgttaa 18
    ||||| ||||| ||||| |||||
Db 575 ATACTTTGGTCATGTAA 558

RESULT 8
AQ150695
LOCUS
DEFINITION
HS_3203_A2_C12_MR CIT Approved Human Genomic Sperm Library D Homo
    sapiens genomic clone Plate-3203 Col-24 Row=E, DNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
human.

```

```

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 404)
AUTHORS
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3203 row: E column: 24
Class: BAC ends
High quality sequence stop: 404.
Location/Qualifiers
1. .404
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3203 Col=24 Row=E"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
    E-Coli DH10B"
BASE COUNT
109 a 87 c 98 g 110 t
ORIGIN

Query Match 88.9%; Score 16; DB 12; Length 404;
Best Local Similarity 100.0%; Pred. No. 6.8e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atactctggtcatgtt 16
    ||||| ||||| ||||| |||||
Db 285 ATACTCGTCATGTT 300

RESULT 9
AW812504/c
LOCUS
DEFINITION
CM4-ST0181-231199-049-b12 ST0181 Homo sapiens cDNA, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 635)
AUTHORS
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

```

Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2=CM4-ST0181-231  
 199-049-b126t3-1999-11-23&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 18  
 High quality sequence stop: 586.

## FEATURES

Location/Qualifiers  
 1..635  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="ST0181"  
 /dev\_stage="Adult"

/note="Organ: stomach; Vector: puc18; Site\_1: SmaI;  
 Site\_2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."  
 203 a 101 c 135 g 196 t

## BASE COUNT

203 a 101 c 135 g 196 t

Query Match 88.9%; Score 16; DB 9; Length 635;  
 Best Local Similarity 100.0%; Pred. No. 7.6e+02;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atactctggtcatgtt 16

Db 94 ATACTCTGGTCATGTT 79

## RESULT 10

CNS02T50/c  
 LOCUS  
 DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone  
 163022 of library G from Tetraodon nigroviridis, genomic survey  
 sequence.  
 AL212759  
 AL212759.1 GI:7871578  
 GSS: genome survey sequence.  
 TETRAODON NIGROVIRIDIS  
 Tetraodon nigroviridis  
 Tetraodon nigroviridis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;  
 Tetraodontidae; Tetraodon.  
 1 (bases 1 to 1026)  
 Roest-Crollius H., Jaillon O., Dasilva C., Fizes C., Fisher C.,  
 Bouneau L., Billault A., Quetier F., Saurin W., Bernot A. and  
 Weissenbach J.  
 Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetraodon nigroviridis  
 Unpublished

## JOURNAL

2 (bases 1 to 1026)  
 Roest-Crollius H., Jaillon O., Dasilva C., Bouneau L., Fisher C.,  
 Bernot A., Fizes C., Wincker P., Brottier P., Quetier F.,  
 Saurin W. and Weissenbach J.  
 Human gene number estimate provided by genome wide analysis using  
 Tetraodon nigroviridis DNA sequence  
 Unpublished

## REFERENCE

3 (bases 1 to 1026)  
 Genoscope.  
 Direct Submission  
 Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
 This sequence is a single read and was generated as part of a large  
 scale clone-end sequencing project of the Tetraodon nigroviridis  
 genome. For more information, please take a look at  
 http://www.genoscope.cns.fr/tetraodon.

## FEATURES

Location/Qualifiers  
 1..1026

/organism="Tetraodon nigroviridis"  
 /db\_xref="taxon:99883"  
 /clone="163022"  
 /clone\_lib="G"  
 /note="Genoscope sequence ID : COAG163BH11L1-end : T7"  
 325 a 201 c 215 g 274 t 11 others

Query Match 86.7%; Score 15.6; DB 12; Length 1026;  
 Best Local Similarity 93.8%; Pred. No. 1.4e+03;  
 Matches 15; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 actctggtcatgttaa 18

Db 497 ACTCTGGTCATKTTAA 482

## RESULT 11

BI287040/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

1 (bases 1 to 213)

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Genome Res. 6 (9), 791-806 (1996)

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dT track served to identify it as a clone from the

non-normalized rat brain pool library cDNA Library Preparation:

M.B. Soares Lab Clone distribution: clones will be available

through Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA-Yes.

Location/Qualifiers

1..213

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clone="UI-R-CT0s-cax-d-05-0-UI"

/clone\_lib="UI-R-CT0s"

/dev\_stage="ADULT"

/lab\_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CT0s

library is a non-normalized library constructed from the

following rat brain tissues: embryonic day 17, embryonic

day 19, embryonic day 21, adult day 1, adult day 12, adult

day 75, adult day 200. For a detailed description of the

library from which this clone was derived, please visit

our web site at rateat.eng.uiowa.edu. The subtraction has

been previously described in (Bonaldo, Lennon and Soares,

Genome Research 6:791-806, 1996)

Query Match 85.6%; Score 15.4; DB 10; Length 213;  
 Best Local Similarity 94.1%; Pred. No. 1.1e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 TAG\_LIB=UI-R-CT0s  
 TAG\_TISSUE=rat brain pool  
 TAG\_SEQ=ACATC  
 BASE COUNT 67 a 35 c 49 g 62 t  
 ORIGIN

2 tactctgtgcatgttaa 18  
 |||||  
 49 TACTCTGGTCATCTTAA 33

## RESULT 12

BE142340 225 bp mRNA linear EST 21-JUN-2000  
 DEFINITION CM3-HT0143-220999-015-b05 HT0143 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BE142340  
 VERSION BE142340.1 GI:8605061  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 225)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bal, G.S., Simpson, D.H.,  
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## MEDLINE

## COMMENT

20202663  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prudente Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=62=CM3-HT0143-220  
 999-015-b05&tl3=1999-09-22&tl4=1)

Seq primer: puc 18 forward

High quality sequence start: 102

High quality sequence stop: 225.

## FEATURES

## source

1. .225  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="HT0143"  
 /dev\_stage="Adult"  
 /note="Organ: head\_neck; Vector: puc18; Site: 1: Smal;  
 Site: 2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the puc 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."  
 52 a 54 c 55 g 64 t

BASE COUNT 52 a 54 c 55 g 64 t  
 ORIGIN

## Query Match

Best Local Similarity 85.6%; Score 15.4; DB 9; Length 225;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggtcatgttta 17  
 |||||  
 Db 76 ATACTCTGGTCATCTTAA 92

## RESULT 13

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

## JOURNAL

## COMMENT

BB024447 297 bp mRNA linear EST 23-JUN-2000  
 BB024447 RIKEN full-length enriched, adult male pituitary gland Mus  
 musculus cDNA clone 5330421N15 3', mRNA sequence.

BB024447 1 GI:8197910

EST.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 297)

Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci

, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,

Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,

Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,

Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.

, Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.

, Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata

, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.

, Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya

, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yananaka, I.

, Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino

, M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)

Unpublished (2000)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S., Sasaki

, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Thermotabilization and thermoactivation of the synthesis of full length

cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)

Itoh, M., Kitzunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,

Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki

, Y. and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation

system. Genome Res. 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

19-44 (1999)

Please visit our web site (http://genome.rtc.riken.go.jp) for

further details.

Location/Qualifiers

1. .297

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="5330421N15"

/tissue="pituitary gland"

/sex="male"

/tissue\_type="pituitary gland"

/dev\_stage="adult"

/lab\_host="DH10B"

/note="Site: 1: SalI; Site: 2: BamHI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in



RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGAGAGAGAGAGCTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 3.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAGAGAGAGAGAGCTCTTTTCTTTTNN 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."

91 a 74 c 43 g 89 t  
BASE COUNT  
ORIGIN

Query Match 85.6%; Score 15.4; DB 9; Length 297;  
Best Local Similarity 94.1%; Pred. No. 1.2e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggtcagtta 17  
||||| |||||||  
Db 172 ATACTCTGATCATGTGA 188

RESULT 14  
BB289468/c  
LOCUS  
DEFINITION  
BB289468  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BB289468 303 bp mRNA linear EST 09-JUL-2000  
clone B020040F24 3', mRNA sequence.  
BB289468  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 303)  
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci  
P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,  
Hirozane,T., Hori,Y., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,  
Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,  
Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,  
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,  
Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata  
F., Shigemoto,Y., Shingawa,A., Shiraki,F., Sogabe,Y., Sugahara,Y.,  
Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya  
T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.,  
Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino  
M., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Konno,H., et al.)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp/  
URL:http://genome.gsc.riken.go.jp/  
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki  
N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Thermotabilization and thermoactivation of thermolabile enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh,M., Katsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,  
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki  
Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.

High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES  
source  
Location/Qualifiers  
1. 303  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="B020040F24"  
/clone\_lib="RIKEN full-length enriched, 2 cells egg"  
/tissue\_type="egg"  
/dev\_stage="2 cells"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGAGAGAGAGAGCTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAGAGAGAGAGAGCTCTTTTCTTTTNN 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 85 a 65 c 70 g 83 t  
ORIGIN

Query Match 85.6%; Score 15.4; DB 9; Length 303;  
Best Local Similarity 94.1%; Pred. No. 1.3e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggtcagtta 17  
||||| |||||||  
Db 87 ATACTCAGTCTCATGTGA 71

RESULT 15  
BF147142  
LOCUS  
DEFINITION  
BF147142  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BF147142 337 bp mRNA linear EST 29-DEC-2000  
uy07f02.y1 McCarrey Eddy spermatocytes Mus musculus cDNA clone  
IMAGE:3657339 5', mRNA sequence.  
BF147142  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 337)  
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person  
B., Waller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter  
E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,  
Waterston,R. and Wilson,R.  
The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)  
Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI:1419643  
Seq primer: Primer name ambiguous  
High quality sequence stop: 333.  
Location/Qualifiers

FEATURES  
source  
Location/Qualifiers  
1. 303  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="B020040F24"  
/clone\_lib="RIKEN full-length enriched, 2 cells egg"  
/tissue\_type="egg"  
/dev\_stage="2 cells"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGAGAGAGAGAGCTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAGAGAGAGAGAGCTCTTTTCTTTTNN 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 85 a 65 c 70 g 83 t  
ORIGIN

Query Match 85.6%; Score 15.4; DB 9; Length 303;  
Best Local Similarity 94.1%; Pred. No. 1.3e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggtcagtta 17  
||||| |||||||  
Db 87 ATACTCAGTCTCATGTGA 71

RESULT 15  
BF147142  
LOCUS  
DEFINITION  
BF147142  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BF147142 337 bp mRNA linear EST 29-DEC-2000  
uy07f02.y1 McCarrey Eddy spermatocytes Mus musculus cDNA clone  
IMAGE:3657339 5', mRNA sequence.  
BF147142  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 337)  
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person  
B., Waller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter  
E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,  
Waterston,R. and Wilson,R.  
The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)  
Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI:1419643  
Seq primer: Primer name ambiguous  
High quality sequence stop: 333.  
Location/Qualifiers

FEATURES  
source  
Location/Qualifiers  
1. 303  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="B020040F24"  
/clone\_lib="RIKEN full-length enriched, 2 cells egg"  
/tissue\_type="egg"  
/dev\_stage="2 cells"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGAGAGAGAGAGCTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAGAGAGAGAGAGCTCTTTTCTTTTNN 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 85 a 65 c 70 g 83 t  
ORIGIN

Query Match 85.6%; Score 15.4; DB 9; Length 303;  
Best Local Similarity 94.1%; Pred. No. 1.3e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggtcagtta 17  
||||| |||||||  
Db 87 ATACTCAGTCTCATGTGA 71

RESULT 15  
BF147142  
LOCUS  
DEFINITION  
BF147142  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BF147142 337 bp mRNA linear EST 29-DEC-2000  
uy07f02.y1 McCarrey Eddy spermatocytes Mus musculus cDNA clone  
IMAGE:3657339 5', mRNA sequence.  
BF147142  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 337)  
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person  
B., Waller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter  
E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,  
Waterston,R. and Wilson,R.  
The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)  
Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI:1419643  
Seq primer: Primer name ambiguous  
High quality sequence stop: 333.  
Location/Qualifiers

FEATURES  
source  
Location/Qualifiers  
1. 303  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="B020040F24"  
/clone\_lib="RIKEN full-length enriched, 2 cells egg"  
/tissue\_type="egg"  
/dev\_stage="2 cells"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGAGAGAGAGAGCTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAGAGAGAGAGAGCTCTTTTCTTTTNN 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 85 a 65 c 70 g 83 t  
ORIGIN

Query Match 85.6%; Score 15.4; DB 9; Length 303;  
Best Local Similarity 94.1%; Pred. No. 1.3e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggtcagtta 17  
||||| |||||||  
Db 87 ATACTCAGTCTCATGTGA 71

RESULT 15  
BF147142  
LOCUS  
DEFINITION  
BF147142  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BF147142 337 bp mRNA linear EST 29-DEC-2000  
uy07f02.y1 McCarrey Eddy spermatocytes Mus musculus cDNA clone  
IMAGE:3657339 5', mRNA sequence.  
BF147142  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 337)  
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person  
B., Waller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter  
E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,  
Waterston,R. and Wilson,R.  
The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)  
Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI:1419643  
Seq primer: Primer name ambiguous  
High quality sequence stop: 333.  
Location/Qualifiers

FEATURES  
source  
Location/Qualifiers  
1. 303  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="B020040F24"  
/clone\_lib="RIKEN full-length enriched, 2 cells egg"  
/tissue\_type="egg"  
/dev\_stage="2 cells"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGAGAGAGAGAGCTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAGAGAGAGAGAGCTCTTTTCTTTTNN 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

```

1. .337
/organism="Mus musculus"
/strain="CD-1"
/db_xref="taxon:10090"
/clone="IWAG:365739"
/clone_lib="McCarrey Eddy spermatocytes"
/sex="male"
/tissue_type="spermatocytes, pooled from multiple mice"
/dev_stage="60 day"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: testis; Vector: pBluescript SK+ (Stratagene
): Site_1: XhoI; Site_2: EcoRI; cDNA oligo dr-primed
[5'-(GTA)10-ACTAGTCTCCAGTGTGTGTGTGT-3'] and directionally
cloned using 5' linkers 5'-AATTGGCAGCAG-3' and
5'-CTCTGCGC-3'. Size selection of >400bp material gives
average insert size ranging from 1-2 kb. Library was mass
excised (from lambda-UniZAP-XR) and resulting
single-stranded phagemids were prepped and transformed
into DH10B. Library contains 98% recombinants.
References: J. Androl. 20:635-639 and Gene 25:263-269.
Library constructed and donated by J. McCarrey, Ph.D.
(Southwest Foundation for Biomedical Research, Dept. of
Genetics); excision done by E.M. Eddy, Ph.D. (National
Institutes of Health, National Institute of Environmental
Health Sciences). Original lambda-based library is
available through ATCC, catalog #63422."
74 a      84 c      60 g      119 t

```

```

Query Match      85.6%; Score 15.4; DB 10; Length 337;
Best local Similarity 94.1%; Pred. No. 1.3e+03;
Matches 16; Conservative 1; Mismatches 0; Gaps 0;

QY      1 atactcgtgctcatgta 17
        ||| ||||| ||| |||
Db      264 ATACTCTGGTCATTTA 280

```

Search completed: July 31, 2002, 12:12:16  
Job time: 13489 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:11:25 ; Search time 720.8 Seconds  
(without alignments)  
42.875 Million cell updates/sec

Title: US-09-899-718A-8  
Perfect score: 18  
Sequence: 1 atactctggcatgttaa 18

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description                  |
|------------|-------|-------------|--------|----|------------------------------|
| C 1        | 16.4  | 91.1        | 32167  | 22 | AK90553 Human digestive sy   |
| C 2        | 16.4  | 91.1        | 44848  | 21 | AA75080 Nucleotide sequenc   |
| C 3        | 15.4  | 85.6        | 936    | 22 | AAK68511 Human immune/haema  |
| C 4        | 14.8  | 82.2        | 407    | 22 | AAK57354 Human immune/haema  |
| C 5        | 14.8  | 82.2        | 474    | 22 | AAK66807 Human immune/haema  |
| C 6        | 14.8  | 82.2        | 474    | 22 | AAK66809 Human immune/haema  |
| C 7        | 14.8  | 82.2        | 596    | 21 | AAAF10198 Fusarium venenatum |
| C 8        | 14.8  | 82.2        | 1327   | 21 | AAAC33932 Arabidopsis thalia |
| C 9        | 14.8  | 82.2        | 1339   | 21 | AAAC34572 Arabidopsis thalia |

|      |      |      |       |    |                              |
|------|------|------|-------|----|------------------------------|
| 10   | 14.8 | 82.2 | 1382  | 21 | AAAC32880 Arabidopsis thalia |
| 11   | 14.8 | 82.2 | 1554  | 21 | AAAC45097 Arabidopsis thalia |
| C 12 | 14.8 | 82.2 | 1926  | 23 | ABL25967 Drosophila melanog  |
| 13   | 14.8 | 82.2 | 1935  | 20 | AAZ30413 Arabidopsis thalia  |
| 14   | 14.8 | 82.2 | 1935  | 22 | AAAF24587 Nucleotide sequenc |
| 15   | 14.8 | 82.2 | 1943  | 22 | AAK67740 Human immune/haema  |
| 16   | 14.8 | 82.2 | 1943  | 22 | AAK67743 Human immune/haema  |
| 17   | 14.8 | 82.2 | 2148  | 21 | AAAC44660 Arabidopsis thalia |
| C 18 | 14.8 | 82.2 | 2339  | 23 | ABL16860 Arabidopsis thalia  |
| C 19 | 14.8 | 82.2 | 2904  | 22 | AAK70130 Drosophila melanog  |
| 20   | 14.8 | 82.2 | 3926  | 23 | ABL25966 Human immune/haema  |
| 21   | 14.8 | 82.2 | 4104  | 23 | ABL25964 Drosophila melanog  |
| C 22 | 14.8 | 82.2 | 5320  | 22 | AAK70131 Drosophila melanog  |
| 23   | 14.8 | 82.2 | 15389 | 22 | AAK80921 Human immune/haema  |
| C 24 | 14.8 | 82.2 | 15610 | 22 | AAK80920 Human reproductive  |
| 25   | 14.8 | 82.2 | 15610 | 22 | AAK80920 Human reproductive  |
| 26   | 14.8 | 82.2 | 39325 | 22 | AAK81660 Human immune/haema  |
| 27   | 14.4 | 80.0 | 614   | 21 | AAK08827 Human immune/haema  |
| 28   | 14.4 | 80.0 | 771   | 22 | AAH07778 Human secreted pro  |
| 29   | 14.4 | 80.0 | 777   | 21 | AAH07778 Human CDNA clone (  |
| C 30 | 14.4 | 80.0 | 855   | 20 | AAZ02318 Human colon cancer  |
| 31   | 14.4 | 80.0 | 888   | 19 | AAV32009 ZOOA protein with   |
| C 32 | 14.4 | 80.0 | 1408  | 19 | AAV49254 Human Rab protein   |
| 33   | 14.4 | 80.0 | 1410  | 16 | AAQ75926 Mouse kappa optate  |
| 34   | 14.4 | 80.0 | 1551  | 22 | AAH14958 Human CDNA sequenc  |
| 35   | 14.4 | 80.0 | 1717  | 21 | AAK39158 Arabidopsis thalia  |
| 36   | 14.4 | 80.0 | 2134  | 23 | AAK76101 DNA encoding novel  |
| C 37 | 14.4 | 80.0 | 2486  | 21 | AAZ89363 Murine mGBP-3 CDNA  |
| C 38 | 14.4 | 80.0 | 3345  | 23 | ABL13550 Drosophila melanog  |
| 39   | 14.4 | 80.0 | 4073  | 23 | ABL08028 Drosophila melanog  |
| 40   | 14.4 | 80.0 | 4123  | 16 | AAQ91282 T. longibrachiatum  |
| C 41 | 14.4 | 80.0 | 4123  | 17 | AAK32221 Trichoderma cellob  |
| C 42 | 14.4 | 80.0 | 4124  | 21 | AAAF21179 Human low adenosin |
| C 43 | 14.4 | 80.0 | 4124  | 21 | AAAF21179 Human low adenosin |
| C 44 | 14.4 | 80.0 | 4729  | 11 | AAQ04043 Recombinant DNA fo  |
| C 45 | 14.4 | 80.0 | 4740  | 21 | AAAF21154 Human low adenosin |

ALIGNMENTS

RESULT 1  
AK90553/c  
ID AK90553 standard; DNA; 32167 BP.

XX AK90553;

XX 05-NOV-2001 (first entry)

XX Human digestive system antigen genomic sequence SEQ ID NO: 4129.

DE Human; digestive system antigen; gene therapy; cancer; appendicitis;

KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;

KW digestive system disorder; Meckel's diverticulum; ds.

XX Homo sapiens.

OS WO200155314-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01324.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

us-09-899-718a-8.rng

Thu Aug 1 08:30:07 2002

PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 05-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 06-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-502630/55.  
XX Polynucleotides encoding digestive system antigens, useful for  
PT diagnosing, treating, preventing and/or prognosing disorders of the  
PT digestive system, particularly cancer and cancer metastases -  
XX Disclosure; SEQ ID NO 4129; 986pp; English.  
PS The present invention provides the protein and coding sequences of a  
XX number of human digestive system antigens. These can be used in the  
CC diagnosis, treatment and prevention of digestive system disorders,  
CC including cancer, Meckel's diverticulum, bacterial or parasitic  
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
CC ulcerative colitis. The present sequence is a genomic DNA fragment  
CC encoding a digestive system antigen of the invention.



PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231245.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0231415.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234224.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246533.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.

PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Barash SC, Ruben SM;  
 PI WPI; 2001-483426/52.  
 XX  
 DR Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis.  
 XX  
 PS Disclosure; SEQ ID NO 23323; 3071pp + Sequence Listing; English.  
 XX  
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patients own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting the  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/haematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/haematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
 CC represent sequences used in the exemplification of the present invention.  
 XX Sequence 936 BP; 265 A; 199 C; 177 G; 295 T; 0 other;  
 SQ  
 Query Match 85.6%; Score 15.4; DB 22; Length 936;  
 Best Local Similarity 94.1%; Pred. No. 1.3e-02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 atactctggctcgttga 17  
 |||||  
 Db 244 ATACTCTGGTTATGTTA 228  
 RESULT 4  
 AAK57354/c  
 ID AAK57354 standard; cDNA; 407 BP.  
 XX  
 AC AAK57354;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:2414.  
 DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
 KW cytostatic; gene therapy; vaccine; metastasis; ss.

|    |                              |              |                 |
|----|------------------------------|--------------|-----------------|
| XX | Homo sapiens.                | 26-SEP-2000; | 2000US-0235484. |
| OS |                              | 27-SEP-2000; | 2000US-0235834. |
| XX |                              | 27-SEP-2000; | 2000US-0235836. |
| PN | WO200157182-A2.              | 29-SEP-2000; | 2000US-0236327. |
| XX |                              | 29-SEP-2000; | 2000US-0236367. |
| XX |                              | 29-SEP-2000; | 2000US-0236368. |
| PD |                              | 29-SEP-2000; | 2000US-0236369. |
| XX | 09-AUG-2001.                 | 29-SEP-2000; | 2000US-0236370. |
| XX |                              | 02-OCT-2000; | 2000US-0236802. |
| XX | 17-JAN-2001; 2001WO-US01354. | 02-OCT-2000; | 2000US-0237037. |
| XX |                              | 02-OCT-2000; | 2000US-0237038. |
| PR | 31-JAN-2000; 2000US-0179065. | 02-OCT-2000; | 2000US-0237039. |
| PR | 04-FEB-2000; 2000US-0180628. | 02-OCT-2000; | 2000US-0237040. |
| PR | 24-FEB-2000; 2000US-0184664. | 13-OCT-2000; | 2000US-0239935. |
| PR | 02-MAR-2000; 2000US-0186350. | 13-OCT-2000; | 2000US-0239937. |
| PR | 16-MAR-2000; 2000US-0189874. | 20-OCT-2000; | 2000US-0240960. |
| PR | 17-MAR-2000; 2000US-0190076. | 20-OCT-2000; | 2000US-0241221. |
| PR | 18-APR-2000; 2000US-0198123. | 20-OCT-2000; | 2000US-0241785. |
| PR | 19-MAY-2000; 2000US-0205515. | 20-OCT-2000; | 2000US-0241786. |
| PR | 07-JUN-2000; 2000US-0209467. | 20-OCT-2000; | 2000US-0241808. |
| PR | 28-JUN-2000; 2000US-0214886. | 20-OCT-2000; | 2000US-0241809. |
| PR | 30-JUN-2000; 2000US-0215135. | 20-OCT-2000; | 2000US-0241826. |
| PR | 07-JUL-2000; 2000US-0216647. | 01-NOV-2000; | 2000US-0244617. |
| PR | 07-JUL-2000; 2000US-0216880. | 08-NOV-2000; | 2000US-0246474. |
| PR | 11-JUL-2000; 2000US-0217487. | 08-NOV-2000; | 2000US-0246475. |
| PR | 11-JUL-2000; 2000US-0217496. | 08-NOV-2000; | 2000US-0246476. |
| PR | 14-JUL-2000; 2000US-0218290. | 08-NOV-2000; | 2000US-0246477. |
| PR | 26-JUL-2000; 2000US-0220963. | 08-NOV-2000; | 2000US-0246478. |
| PR | 26-JUL-2000; 2000US-0220964. | 08-NOV-2000; | 2000US-0246523. |
| PR | 14-AUG-2000; 2000US-0224518. | 08-NOV-2000; | 2000US-0246524. |
| PR | 14-AUG-2000; 2000US-0224519. | 08-NOV-2000; | 2000US-0246525. |
| PR | 14-AUG-2000; 2000US-0225213. | 08-NOV-2000; | 2000US-0246526. |
| PR | 14-AUG-2000; 2000US-0225214. | 08-NOV-2000; | 2000US-0246527. |
| PR | 14-AUG-2000; 2000US-0225266. | 08-NOV-2000; | 2000US-0246528. |
| PR | 14-AUG-2000; 2000US-0225267. | 08-NOV-2000; | 2000US-0246532. |
| PR | 14-AUG-2000; 2000US-0225268. | 08-NOV-2000; | 2000US-0246609. |
| PR | 14-AUG-2000; 2000US-0225270. | 08-NOV-2000; | 2000US-0246610. |
| PR | 14-AUG-2000; 2000US-0225447. | 08-NOV-2000; | 2000US-0246611. |
| PR | 14-AUG-2000; 2000US-0225757. | 08-NOV-2000; | 2000US-0246613. |
| PR | 14-AUG-2000; 2000US-0225758. | 17-NOV-2000; | 2000US-0249207. |
| PR | 14-AUG-2000; 2000US-0225759. | 17-NOV-2000; | 2000US-0249208. |
| PR | 18-AUG-2000; 2000US-0226279. | 17-NOV-2000; | 2000US-0249209. |
| PR | 22-AUG-2000; 2000US-0226681. | 17-NOV-2000; | 2000US-0249210. |
| PR | 22-AUG-2000; 2000US-0226688. | 17-NOV-2000; | 2000US-0249211. |
| PR | 22-AUG-2000; 2000US-0227182. | 17-NOV-2000; | 2000US-0249212. |
| PR | 23-AUG-2000; 2000US-0227182. | 17-NOV-2000; | 2000US-0249213. |
| PR | 30-AUG-2000; 2000US-0228924. | 17-NOV-2000; | 2000US-0249214. |
| PR | 01-SEP-2000; 2000US-0229287. | 17-NOV-2000; | 2000US-0249215. |
| PR | 01-SEP-2000; 2000US-0229343. | 17-NOV-2000; | 2000US-0249216. |
| PR | 01-SEP-2000; 2000US-0229344. | 17-NOV-2000; | 2000US-0249217. |
| PR | 01-SEP-2000; 2000US-0229345. | 17-NOV-2000; | 2000US-0249218. |
| PR | 05-SEP-2000; 2000US-0229509. | 17-NOV-2000; | 2000US-0249219. |
| PR | 05-SEP-2000; 2000US-0229513. | 17-NOV-2000; | 2000US-0249244. |
| PR | 06-SEP-2000; 2000US-0230437. | 17-NOV-2000; | 2000US-0249245. |
| PR | 06-SEP-2000; 2000US-0230438. | 17-NOV-2000; | 2000US-0249264. |
| PR | 08-SEP-2000; 2000US-0231242. | 17-NOV-2000; | 2000US-0249265. |
| PR | 08-SEP-2000; 2000US-0231243. | 17-NOV-2000; | 2000US-0249297. |
| PR | 08-SEP-2000; 2000US-0231244. | 17-NOV-2000; | 2000US-0249299. |
| PR | 08-SEP-2000; 2000US-0231413. | 17-NOV-2000; | 2000US-0249300. |
| PR | 08-SEP-2000; 2000US-0231414. | 01-DEC-2000; | 2000US-0250160. |
| PR | 08-SEP-2000; 2000US-0232080. | 01-DEC-2000; | 2000US-0250391. |
| PR | 08-SEP-2000; 2000US-0232081. | 05-DEC-2000; | 2000US-0251030. |
| PR | 12-SEP-2000; 2000US-0231968. | 05-DEC-2000; | 2000US-0251988. |
| PR | 14-SEP-2000; 2000US-0232397. | 06-DEC-2000; | 2000US-0251479. |
| PR | 14-SEP-2000; 2000US-0232398. | 06-DEC-2000; | 2000US-0251856. |
| PR | 14-SEP-2000; 2000US-0232399. | 08-DEC-2000; | 2000US-0251869. |
| PR | 14-SEP-2000; 2000US-0232400. | 08-DEC-2000; | 2000US-0251989. |
| PR | 14-SEP-2000; 2000US-0232401. | 08-DEC-2000; | 2000US-0251990. |
| PR | 14-SEP-2000; 2000US-0233063. | 11-DEC-2000; | 2000US-0254097. |
| PR | 14-SEP-2000; 2000US-0233064. | 05-JAN-2001; | 2001US-0259678. |
| PR | 14-SEP-2000; 2000US-0233065. |              |                 |
| PR | 21-SEP-2000; 2000US-0233065. |              |                 |
| PR | 21-SEP-2000; 2000US-0234223. |              |                 |
| PR | 21-SEP-2000; 2000US-0234274. |              |                 |
| PR | 25-SEP-2000; 2000US-0234997. |              |                 |
| PR | 25-SEP-2000; 2000US-0234998. |              |                 |
| PR |                              |              | XX              |

PA (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Barash SC, Ruben SM;  
 PI  
 XX  
 XX  
 DR WPI; 2001-483426/52.  
 DR P-PSDB; AAK84573.  
 XX  
 PT Nucleic acids encoding human immune/haematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis -  
 PT  
 XX  
 XX Claim 1; SEQ ID NO 2414; 3071pp + Sequence Listing; English.  
 PS  
 XX  
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/haematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/haematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
 CC represent sequences used in the exemplification of the present invention.  
 XX  
 XX Sequence 407 BP; 105 A; 102 C; 56 G; 140 T; 4 other;  
 SQ

Query Match 82.2%; Score 14.8; DB 22; Length 407;  
 Best Local Similarity 88.9%; Pred No. 2.4e+02;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atactctggtcatgtaa 18  
 ||| ||||| |||||  
 Db 133 ATAGTCTGGTAATGTTAA 116

RESULT 5  
 AAK66807  
 ID AAK66807 standard; DNA; 474 BP.  
 XX  
 AC AAK66807;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:21619.  
 XX  
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
 KW cytostatic; gene therapy; vaccine; metastasis; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157182-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01354.  
 XX  
 XX 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR

PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226861.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240980.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.



PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
PR XX XX  
PR (HUMA-) HUMAN GENOME SCI INC.  
PR  
PR Rosen CA, Barash SC, Ruben SM;  
PR  
PR WPT; 2001-483426/52.  
PR  
PR Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PR useful for preventing, diagnosing and/or treating cancers and  
PR metastasis -  
PR  
PR  
PR Disclosure; SEQ ID NO 21619; 307lpp + Sequence Listing; English.  
PR  
PR AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
PR amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
PR activity, and can be used in gene therapy and vaccine production. (I)  
PR proteins and polynucleotides may be used in the prevention, diagnosis and  
PR treatment of diseases associated with inappropriate (I) expression. For  
PR example, they may be used to treat disorders associated with decreased

CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention.  
CC  
XX  
SQ Sequence 474 BP; 135 A; 109 C; 90 G; 140 T; 0 other;

Query Match 82.2%; Score 14.8; DB 22; Length 474;

Best Local Similarity 88.9%; Pred. No. 2.5e+02;

Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 atactctgtcatgttaa 18  
||| ||||| |||||  
Db 392 ataactgtcgaagttaa 409

#### RESULT 6

AAK66809

ID AAK66809 standard; DNA; 474 BP.

XX AAK66809;

AC AAK66809;

XX 06-NOV-2001 (first entry)

DT 06-NOV-2001 (first entry)

XX Human immune/hematopoietic antigen genomic sequence SEQ ID NO:21621.

DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KW cytostatic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

OS WO200157182-A2.

XX 09-AUG-2001.

PD 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

|    |              |                 |
|----|--------------|-----------------|
| PR | 14-AUG-2000; | 2000US-0255758; |
| PR | 14-AUG-2000; | 2000US-0257519; |
| PR | 18-AUG-2000; | 2000US-0226279; |
| PR | 22-AUG-2000; | 2000US-0226681; |
| PR | 22-AUG-2000; | 2000US-0226686; |
| PR | 22-AUG-2000; | 2000US-0227182; |
| PR | 23-AUG-2000; | 2000US-0227009; |
| PR | 30-AUG-2000; | 2000US-0228924; |
| PR | 01-SEP-2000; | 2000US-0229287; |
| PR | 01-SEP-2000; | 2000US-0229343; |
| PR | 01-SEP-2000; | 2000US-0229344; |
| PR | 01-SEP-2000; | 2000US-0229345; |
| PR | 05-SEP-2000; | 2000US-0229509; |
| PR | 05-SEP-2000; | 2000US-0229513; |
| PR | 06-SEP-2000; | 2000US-0230437; |
| PR | 06-SEP-2000; | 2000US-0230438; |
| PR | 08-SEP-2000; | 2000US-0231242; |
| PR | 08-SEP-2000; | 2000US-0231243; |
| PR | 08-SEP-2000; | 2000US-0231244; |
| PR | 08-SEP-2000; | 2000US-0231413; |
| PR | 08-SEP-2000; | 2000US-0231414; |
| PR | 08-SEP-2000; | 2000US-0232080; |
| PR | 08-SEP-2000; | 2000US-0232081; |
| PR | 12-SEP-2000; | 2000US-0231968; |
| PR | 14-SEP-2000; | 2000US-0232397; |
| PR | 14-SEP-2000; | 2000US-0232398; |
| PR | 14-SEP-2000; | 2000US-0232399; |
| PR | 21-SEP-2000; | 2000US-0234223; |
| PR | 21-SEP-2000; | 2000US-0234274; |
| PR | 25-SEP-2000; | 2000US-0234997; |
| PR | 25-SEP-2000; | 2000US-0234998; |
| PR | 26-SEP-2000; | 2000US-0235834; |
| PR | 27-SEP-2000; | 2000US-0235834; |
| PR | 27-SEP-2000; | 2000US-0235836; |
| PR | 29-SEP-2000; | 2000US-0236327; |
| PR | 29-SEP-2000; | 2000US-0236327; |
| PR | 29-SEP-2000; | 2000US-0236367; |
| PR | 29-SEP-2000; | 2000US-0236368; |
| PR | 29-SEP-2000; | 2000US-0236369; |
| PR | 29-SEP-2000; | 2000US-0236370; |
| PR | 02-OCT-2000; | 2000US-0236802; |
| PR | 02-OCT-2000; | 2000US-0236802; |
| PR | 02-OCT-2000; | 2000US-0237038; |
| PR | 02-OCT-2000; | 2000US-0237039; |
| PR | 13-OCT-2000; | 2000US-0237040; |
| PR | 13-OCT-2000; | 2000US-0239935; |
| PR | 13-OCT-2000; | 2000US-0246477; |
| PR | 20-OCT-2000; | 2000US-0240960; |
| PR | 20-OCT-2000; | 2000US-0241221; |
| PR | 20-OCT-2000; | 2000US-0241175; |
| PR | 20-OCT-2000; | 2000US-0244617; |
| PR | 20-OCT-2000; | 2000US-0241786; |
| PR | 20-OCT-2000; | 2000US-0241787; |
| PR | 20-OCT-2000; | 2000US-0241808; |
| PR | 20-OCT-2000; | 2000US-0241809; |
| PR | 20-OCT-2000; | 2000US-0241826; |
| PR | 01-NOV-2000; | 2000US-0244617; |
| PR | 08-NOV-2000; | 2000US-0246478; |
| PR | 08-NOV-2000; | 2000US-0246475; |
| PR | 08-NOV-2000; | 2000US-0246524; |
| PR | 08-NOV-2000; | 2000US-0246525; |
| PR | 08-NOV-2000; | 2000US-0248526; |
| PR | 08-NOV-2000; | 2000US-0248527; |
| PR | 08-NOV-2000; | 2000US-0248528; |
| PR | 08-NOV-2000; | 2000US-0248532; |
| PR | 08-NOV-2000; | 2000US-0246609; |

|     |                                    |                 |
|-----|------------------------------------|-----------------|
| PR  | 08-NOV-2000;                       | 2000US-0246610. |
| PR  | 08-NOV-2000;                       | 2000US-0246611. |
| PR  | 08-NOV-2000;                       | 2000US-0246613. |
| PR  | 17-NOV-2000;                       | 2000US-0249207. |
| PR  | 17-NOV-2000;                       | 2000US-0249208. |
| PR  | 17-NOV-2000;                       | 2000US-0249209. |
| PR  | 17-NOV-2000;                       | 2000US-0249210. |
| PR  | 17-NOV-2000;                       | 2000US-0249211. |
| PR  | 17-NOV-2000;                       | 2000US-0249212. |
| PR  | 17-NOV-2000;                       | 2000US-0249213. |
| PR  | 17-NOV-2000;                       | 2000US-0249214. |
| PR  | 17-NOV-2000;                       | 2000US-0249215. |
| PR  | 17-NOV-2000;                       | 2000US-0249216. |
| PR  | 17-NOV-2000;                       | 2000US-0249217. |
| PR  | 17-NOV-2000;                       | 2000US-0249218. |
| PR  | 17-NOV-2000;                       | 2000US-0249244. |
| PR  | 17-NOV-2000;                       | 2000US-0249245. |
| PR  | 17-NOV-2000;                       | 2000US-0249264. |
| PR  | 17-NOV-2000;                       | 2000US-0249265. |
| PR  | 17-NOV-2000;                       | 2000US-0249297. |
| PR  | 17-NOV-2000;                       | 2000US-0249299. |
| PR  | 01-DEC-2000;                       | 2000US-0250300. |
| PR  | 01-DEC-2000;                       | 2000US-0250360. |
| PR  | 05-DEC-2000;                       | 2000US-0250391. |
| PR  | 05-DEC-2000;                       | 2000US-0251030. |
| PR  | 05-DEC-2000;                       | 2000US-0251988. |
| PR  | 06-DEC-2000;                       | 2000US-0256719. |
| PR  | 06-DEC-2000;                       | 2000US-0256719. |
| PR  | 08-DEC-2000;                       | 2000US-0251856. |
| PR  | 08-DEC-2000;                       | 2000US-0251868. |
| PR  | 08-DEC-2000;                       | 2000US-0251869. |
| PR  | 08-DEC-2000;                       | 2000US-0251990. |
| PR  | 11-DEC-2000;                       | 2000US-0254097. |
| PR  | 05-JAN-2001;                       | 2001US-0259678. |
| XX  |                                    |                 |
| XX  | (HUMA-) HUMAN GENOME SCI INC       |                 |
| XX  |                                    |                 |
| PIL | Rosen CA, Barash SC, Ruben         |                 |
| XX  | WPI; 2001-483426/52.               |                 |
| DR  | Nucleic acids encoding human       |                 |
| PT  | useful for preventing, diagnosing, |                 |
| PT  | metastasis -                       |                 |
| XX  |                                    |                 |
| FS  | Disclosure; SEQ ID NO 21621;       |                 |
| XX  |                                    |                 |
| CC  | AAK54951 to AAK64702 encode t      |                 |
| CC  | amino acid sequences given in      |                 |
| CC  | activity, and can be used in       |                 |
| CC  | treatments and polynucleotides     |                 |
| CC  | proteins of diseases associat      |                 |
| CC  | example, they may be used to       |                 |
| CC  | expression by rectifying muta      |                 |
| CC  | that affect the activity of        |                 |
| CC  | polynucleotides may be used        |                 |
| CC  | the nucleic acids into a host      |                 |
| CC  | protein. (I) proteins and po       |                 |
| CC  | diagnose and treat immune/hae      |                 |
| CC  | cancers and cancer metastases      |                 |
| CC  | to AAK87694 represent human        |                 |
| CC  | sequences from the present in      |                 |
| CC  | represent sequences used in        |                 |
| XX  |                                    |                 |
| SO  | Sequence 474 BP; 135 A; 109 C      |                 |

|                       |              |                    |               |             |
|-----------------------|--------------|--------------------|---------------|-------------|
| Query Match           | 82.2%;       | Score 14.8;        | DB 22;        | Length 474; |
| Best Local Similarity | 88.9%;       | Pred. No. 2.5e+02; |               |             |
| Matches 16:           | Conservative | 0;                 | Mismatches 2; | Indels 0    |

QY 1 atactctggtcatgttaa 18  
 III |IIIIIIII |IIII  
 Db 392 ataactctggtcaagttaa 409

## RESULT 7

AAF10198  
 ID AAF10198 standard; cDNA; 596 BP.  
 AC AAF10198;  
 XX  
 XX  
 DT 13-MAR-2001 (first entry)  
 XX  
 DE Fusarium venenatum EST SEQ ID NO:2721.  
 XX

XX Multiple gene expression; filamentous fungal cell; EST;  
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
 KW culture condition; environmental stress; spore morphogenesis;  
 KW metabolic pathway engineering; catabolic pathway engineering; ss.  
 XX  
 OS Fusarium venenatum.

XX WO2000056762-A2.  
 PN  
 XX  
 PD 28-SEP-2000.

XX 22-MAR-2000; 2000WO-US07781.  
 XX  
 XX 22-MAR-1999; 99US-0273623.

XX (NOVO ) NOVO NORDISK BIOTECH INC.  
 PA (NOVO ) NOVO NORDISK AS.  
 XX  
 XX

PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
 XX WPI; 2000-594572/56.

XX Monitoring differential expression of genes in filamentous fungal cells  
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
 PT substrate of expressed sequence tags -

XX Claim 86; Page 1372-1373; 3161pp; English.

XX The present invention describes a method for monitoring differential  
 CC expression of genes in a first filamentous fungal (FF) cell relative to  
 CC expression of the same genes in one or more second filamentous fungal  
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 CC are used in the methods for monitoring differential expression of genes  
 CC in a first filamentous fungal (FF) cell relative to expression of the  
 CC same genes in one or more second filamentous fungal cells. Monitoring  
 CC the global expression of genes from FF cells allows the production  
 CC potential of the microorganisms to be improved. New genes may be  
 CC discovered, possible functions of unknown open reading frames can be  
 CC identified and gene copy number variation and stability can be  
 CC monitored. The expression of genes can be used to study how FF cells  
 CC adapt to changes in culture conditions, environmental stress, spore  
 CC morphogenesis, recombination, metabolic or catabolic pathway  
 CC engineering. Using ESTs provides several advantages over genomic or  
 CC random cDNA clones including elimination of redundancy as one spot on an  
 CC array equals one gene or open reading frame, and organisation of the  
 CC microarrays based on function of the gene products to facilitate  
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from  
 CC Fusarium venenatum; AAF11248 to AAF11953 represents ESTs from Aspergillus  
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and  
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are  
 CC all specifically claimed in the present invention.

XX Sequence 596 BP; 156 A; 146 C; 149 G; 139 T; 6 other;

Query Match

82.2%; Score 14.8; DB 21; Length 596;

Best Local Similarity 88.9%; Pred. No. 2.5e+02;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 atactctggtcatgttaa 18  
 IIIIIII |IIIIIIII |  
 Db 565 atactctggtcatgttaa 582

## RESULT 8

AAC33932  
 ID AAC33932 standard; DNA; 1327 BP.

XX AAC33932;  
 AC  
 XX

DT 17-OCT-2000 (first entry)  
 XX

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 4839.

XX Hybridisation assay; genetic mapping; gene expression control;  
 KW protein identification; signal transduction pathway;  
 KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EF1033405-A2.  
 PN  
 XX  
 PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.  
 XX  
 XX 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126284.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144332.  
 PR 20-JUL-1999; 99US-0144637.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 82.2%; Score 14.8; DB 21; Length 1327;  
 Best Local Similarity 88.9%; Pred. No. 2.7e+02;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 atactctggtcatgttaa 18  
 || ||||| |||||  
 Db 269 atccctggttatgttaa 286

RESULT 9  
AAC34572  
ID AAC34572 standard; DNA; 1339 BP.  
XX AC AAC34572;  
XX DT 17-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 7122.  
XX KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126284.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135124.  
PR 24-MAY-1999; 99US-0135353.  
PR 25-MAY-1999; 99US-0135629.  
PR 27-MAY-1999; 99US-0136021.  
PR 28-MAY-1999; 99US-0136382.  
PR 01-JUN-1999; 99US-0136782.  
PR 03-JUN-1999; 99US-0137222.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137502.  
PR 08-JUN-1999; 99US-0137724.  
PR 10-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 27-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148171.  
PR 12-AUG-1999; 99US-0148319.  
PR 13-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.

us-09-899-718a-8.rng

Thu Aug 1 08:30:07 2002

|    |              |               |    |                         |                                           |
|----|--------------|---------------|----|-------------------------|-------------------------------------------|
| PR | 20-AUG-1999; | 99US-0149723. | DT | 17-OCT-2000             | (first entry)                             |
| PR | 20-AUG-1999; | 99US-0149929. | XX | Arabidopsis thaliana    | DNA fragment SEQ ID NO: 983.              |
| PR | 23-AUG-1999; | 99US-0149902. | XX | Hybridisation assay;    | genetic mapping; gene expression control; |
| PR | 23-AUG-1999; | 99US-0149930. | KW | protein identification; | signal transduction pathway;              |
| PR | 25-AUG-1999; | 99US-0150566. | KW | metabolic pathway;      | promoter; termination sequence; ss.       |
| PR | 26-AUG-1999; | 99US-0150884. | XX | Arabidopsis thaliana.   |                                           |
| PR | 27-AUG-1999; | 99US-0151065. | OS |                         |                                           |
| PR | 27-AUG-1999; | 99US-0151066. | XX | EP1033405-A2.           |                                           |
| PR | 27-AUG-1999; | 99US-0151080. | PN | 06-SEP-2000.            |                                           |
| PR | 30-AUG-1999; | 99US-0151303. | XX | 25-FEB-2000;            | 2000EP-0301439.                           |
| PR | 31-AUG-1999; | 99US-0151438. | XX | 25-FEB-1999;            | 99US-0121825.                             |
| PR | 01-SEP-1999; | 99US-0151930. | PR | 05-MAR-1999;            | 99US-0123180.                             |
| PR | 07-SEP-1999; | 99US-0152363. | PR | 09-MAR-1999;            | 99US-0123548.                             |
| PR | 10-SEP-1999; | 99US-0153070. | PR | 23-MAR-1999;            | 99US-0125788.                             |
| PR | 13-SEP-1999; | 99US-0153758. | PR | 25-MAR-1999;            | 99US-0126264.                             |
| PR | 15-SEP-1999; | 99US-0154018. | PR | 29-MAR-1999;            | 99US-0126785.                             |
| PR | 16-SEP-1999; | 99US-0154039. | PR | 01-APR-1999;            | 99US-0127462.                             |
| PR | 20-SEP-1999; | 99US-0154779. | PR | 06-APR-1999;            | 99US-0128234.                             |
| PR | 22-SEP-1999; | 99US-0155139. | PR | 08-APR-1999;            | 99US-0128714.                             |
| PR | 23-SEP-1999; | 99US-0155486. | PR | 16-APR-1999;            | 99US-0128845.                             |
| PR | 24-SEP-1999; | 99US-0155659. | PR | 19-APR-1999;            | 99US-0130077.                             |
| PR | 28-SEP-1999; | 99US-0156458. | PR | 21-APR-1999;            | 99US-0130449.                             |
| PR | 29-SEP-1999; | 99US-0156596. | PR | 23-APR-1999;            | 99US-0130510.                             |
| PR | 04-OCT-1999; | 99US-0157117. | PR | 28-APR-1999;            | 99US-0130891.                             |
| PR | 05-OCT-1999; | 99US-0157753. | PR | 30-APR-1999;            | 99US-0131449.                             |
| PR | 06-OCT-1999; | 99US-0157865. | PR | 04-MAY-1999;            | 99US-0132407.                             |
| PR | 07-OCT-1999; | 99US-0158029. | PR | 05-MAY-1999;            | 99US-0132485.                             |
| PR | 08-OCT-1999; | 99US-0158222. | PR | 06-MAY-1999;            | 99US-0132486.                             |
| PR | 12-OCT-1999; | 99US-0158369. | PR | 07-MAY-1999;            | 99US-0132863.                             |
| PR | 13-OCT-1999; | 99US-0159293. | PR | 11-MAY-1999;            | 99US-0134256.                             |
| PR | 13-OCT-1999; | 99US-0159294. | PR | 14-MAY-1999;            | 99US-0134218.                             |
| PR | 13-OCT-1999; | 99US-0159295. | PR | 14-MAY-1999;            | 99US-0134219.                             |
| PR | 14-OCT-1999; | 99US-0159329. | PR | 14-MAY-1999;            | 99US-0134221.                             |
| PR | 14-OCT-1999; | 99US-0159330. | PR | 14-MAY-1999;            | 99US-0134370.                             |
| PR | 14-OCT-1999; | 99US-0159331. | PR | 18-MAY-1999;            | 99US-0134768.                             |
| PR | 14-OCT-1999; | 99US-0159637. | PR | 19-MAY-1999;            | 99US-0134941.                             |
| PR | 14-OCT-1999; | 99US-0159638. | PR | 20-MAY-1999;            | 99US-0135124.                             |
| PR | 18-OCT-1999; | 99US-0159584. | PR | 21-MAY-1999;            | 99US-0135353.                             |
| PR | 21-OCT-1999; | 99US-0160767. | PR | 24-MAY-1999;            | 99US-0135629.                             |
| PR | 21-OCT-1999; | 99US-0160768. | PR | 25-MAY-1999;            | 99US-0136021.                             |
| PR | 21-OCT-1999; | 99US-0160770. | PR | 27-MAY-1999;            | 99US-0136392.                             |
| PR | 21-OCT-1999; | 99US-0160814. | PR | 28-MAY-1999;            | 99US-0136782.                             |
| PR | 21-OCT-1999; | 99US-0160815. | PR | 01-JUN-1999;            | 99US-0137222.                             |
| PR | 22-OCT-1999; | 99US-0160981. | PR | 03-JUN-1999;            | 99US-0137528.                             |
| PR | 22-OCT-1999; | 99US-0160989. | PR | 04-JUN-1999;            | 99US-0137502.                             |
| PR | 25-OCT-1999; | 99US-0161404. | PR | 07-JUN-1999;            | 99US-0137724.                             |
| PR | 25-OCT-1999; | 99US-0161405. | PR | 08-JUN-1999;            | 99US-0138094.                             |
| PR | 25-OCT-1999; | 99US-0161406. | PR | 10-JUN-1999;            | 99US-0138540.                             |
| PR | 26-OCT-1999; | 99US-0161359. | PR | 10-JUN-1999;            | 99US-0138847.                             |
| PR | 26-OCT-1999; | 99US-0161360. | PR | 14-JUN-1999;            | 99US-0139119.                             |
| PR | 26-OCT-1999; | 99US-0161361. | PR | 16-JUN-1999;            | 99US-0139452.                             |
| PR | 28-OCT-1999; | 99US-0161920. | PR | 16-JUN-1999;            | 99US-0139453.                             |
| PR | 28-OCT-1999; | 99US-0161992. | PR | 17-JUN-1999;            | 99US-0139492.                             |
| PR | 28-OCT-1999; | 99US-0161993. | PR | 18-JUN-1999;            | 99US-0139454.                             |
| PR | 28-OCT-1999; | 99US-0161993. | PR | 18-JUN-1999;            | 99US-0139455.                             |
| PR | 29-OCT-1999; | 99US-0162142. | PR | 18-JUN-1999;            | 99US-0139456.                             |
|    |              |               | PR | 18-JUN-1999;            | 99US-0139457.                             |
|    |              |               | PR | 18-JUN-1999;            | 99US-0139458.                             |
|    |              |               | PR | 18-JUN-1999;            | 99US-0139459.                             |
|    |              |               | PR | 18-JUN-1999;            | 99US-0139460.                             |
|    |              |               | PR | 18-JUN-1999;            | 99US-0139461.                             |
|    |              |               | PR | 18-JUN-1999;            | 99US-0139462.                             |
|    |              |               | PR | 18-JUN-1999;            | 99US-0139463.                             |
|    |              |               | PR | 18-JUN-1999;            | 99US-0139750.                             |

|                       |              |                    |               |              |
|-----------------------|--------------|--------------------|---------------|--------------|
| Query Match           | 82.28;       | Score 14.8;        | DB 21;        | Length 1339; |
| Best Local Similarity | 88.98;       | Pred. No. 2.7e+02; |               |              |
| Matches 16;           | Conservative | 0;                 | Mismatches 2; | Indels 0;    |
| Gaps                  | 0;           |                    |               |              |

|    |     |                   |     |
|----|-----|-------------------|-----|
| QY | 1   | atactctggtcatgtaa | 18  |
| DB | 264 | atcctctggtcatgtaa | 281 |
|    |     |                   |     |

|          |                                  |
|----------|----------------------------------|
| RESULT   | 10                               |
| AAC32880 |                                  |
| ID       | AAC32880 standard; DNA; 1382 BP. |
| XX       |                                  |
| AC       | AAC32880;                        |
| XX       |                                  |

PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139859.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 24-JUN-1999; 99US-0140354.  
 PR 25-JUN-1999; 99US-0140695.  
 PR 26-JUN-1999; 99US-0140823.  
 PR 27-JUN-1999; 99US-0140991.  
 PR 28-JUN-1999; 99US-0141287.  
 PR 29-JUN-1999; 99US-0141842.  
 PR 30-JUN-1999; 99US-0142154.  
 PR 01-JUL-1999; 99US-0142055.  
 PR 02-JUL-1999; 99US-0142390.  
 PR 03-JUL-1999; 99US-0142803.  
 PR 04-JUL-1999; 99US-0142920.  
 PR 05-JUL-1999; 99US-0142977.  
 PR 06-JUL-1999; 99US-0143542.  
 PR 07-JUL-1999; 99US-0143624.  
 PR 08-JUL-1999; 99US-0144005.  
 PR 09-JUL-1999; 99US-0144085.  
 PR 10-JUL-1999; 99US-0144086.  
 PR 11-JUL-1999; 99US-0144322.  
 PR 12-JUL-1999; 99US-0144331.  
 PR 13-JUL-1999; 99US-0144332.  
 PR 14-JUL-1999; 99US-0144333.  
 PR 15-JUL-1999; 99US-0144334.  
 PR 16-JUL-1999; 99US-0144335.  
 PR 17-JUL-1999; 99US-0144332.  
 PR 18-JUL-1999; 99US-0144632.  
 PR 19-JUL-1999; 99US-0144884.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 22-JUL-1999; 99US-0145086.  
 PR 23-JUL-1999; 99US-0145087.  
 PR 24-JUL-1999; 99US-0145087.  
 PR 25-JUL-1999; 99US-0145089.  
 PR 26-JUL-1999; 99US-0145192.  
 PR 27-JUL-1999; 99US-0145145.  
 PR 28-JUL-1999; 99US-0145224.  
 PR 29-JUL-1999; 99US-0145276.  
 PR 30-JUL-1999; 99US-0145913.  
 PR 31-JUL-1999; 99US-0145918.  
 PR 01-AUG-1999; 99US-0145919.  
 PR 02-AUG-1999; 99US-0145951.  
 PR 03-AUG-1999; 99US-0146386.  
 PR 04-AUG-1999; 99US-0146388.  
 PR 05-AUG-1999; 99US-0147038.  
 PR 06-AUG-1999; 99US-0147204.  
 PR 07-AUG-1999; 99US-0147302.  
 PR 08-AUG-1999; 99US-0147192.  
 PR 09-AUG-1999; 99US-0147260.  
 PR 10-AUG-1999; 99US-0147303.  
 PR 11-AUG-1999; 99US-0147416.  
 PR 12-AUG-1999; 99US-0147493.  
 PR 13-AUG-1999; 99US-0147935.  
 PR 14-AUG-1999; 99US-0148171.  
 PR 15-AUG-1999; 99US-0148319.  
 PR 16-AUG-1999; 99US-0148341.  
 PR 17-AUG-1999; 99US-0148565.  
 PR 18-AUG-1999; 99US-0148684.  
 PR 19-AUG-1999; 99US-0149368.  
 PR 20-AUG-1999; 99US-0149175.  
 PR 21-AUG-1999; 99US-0149426.  
 PR 22-AUG-1999; 99US-0149722.  
 PR 23-AUG-1999; 99US-0149723.  
 PR 24-AUG-1999; 99US-0149929.  
 PR 25-AUG-1999; 99US-0149930.  
 PR 26-AUG-1999; 99US-0150566.  
 PR 27-AUG-1999; 99US-0151065.

PR 27-AUG-1999; 99US-0151066.  
 PR 28-AUG-1999; 99US-0151080.  
 PR 29-AUG-1999; 99US-0151303.  
 PR 30-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 02-SEP-1999; 99US-0152363.  
 PR 03-SEP-1999; 99US-0153070.  
 PR 04-SEP-1999; 99US-0153758.  
 PR 05-SEP-1999; 99US-0154039.  
 PR 06-SEP-1999; 99US-0154779.  
 PR 07-SEP-1999; 99US-0155139.  
 PR 08-SEP-1999; 99US-0155486.  
 PR 09-SEP-1999; 99US-0155659.  
 PR 10-SEP-1999; 99US-0156458.  
 PR 11-SEP-1999; 99US-0156596.  
 PR 12-SEP-1999; 99US-0157117.  
 PR 13-SEP-1999; 99US-0157753.  
 PR 14-SEP-1999; 99US-0157865.  
 PR 15-SEP-1999; 99US-0158029.  
 PR 16-SEP-1999; 99US-0158232.  
 PR 17-SEP-1999; 99US-0158369.  
 PR 18-SEP-1999; 99US-0159293.  
 PR 19-SEP-1999; 99US-0159294.  
 PR 20-SEP-1999; 99US-0159295.  
 PR 21-SEP-1999; 99US-0159329.  
 PR 22-SEP-1999; 99US-0159330.  
 PR 23-SEP-1999; 99US-0159331.  
 PR 24-SEP-1999; 99US-0159637.  
 PR 25-SEP-1999; 99US-0159638.  
 PR 26-SEP-1999; 99US-0159584.  
 PR 27-SEP-1999; 99US-0160741.  
 PR 28-SEP-1999; 99US-0160767.  
 PR 29-SEP-1999; 99US-0160770.  
 PR 30-SEP-1999; 99US-0160770.  
 PR 01-OCT-1999; 99US-0160815.  
 PR 02-OCT-1999; 99US-0160981.  
 PR 03-OCT-1999; 99US-0160989.  
 PR 04-OCT-1999; 99US-0161404.  
 PR 05-OCT-1999; 99US-0161405.  
 PR 06-OCT-1999; 99US-0161406.  
 PR 07-OCT-1999; 99US-0161359.  
 PR 08-OCT-1999; 99US-0161360.  
 PR 09-OCT-1999; 99US-0161361.  
 PR 10-OCT-1999; 99US-0161361.  
 PR 11-OCT-1999; 99US-0161920.  
 PR 12-OCT-1999; 99US-0161992.  
 PR 13-OCT-1999; 99US-0161993.  
 PR 14-OCT-1999; 99US-0162142.

Query Match 82.2%; Score 14.8; DB 21; Length 1382;  
 Best Local Similarity 88.9%; Pred. No. 2.7e+02;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atactctggtcatgttaa 18  
 Db 349 atctctggtcatgttaa 366

RESULT 11  
 AAC45097  
 ID AAC45097 standard; DNA; 1554 BP.  
 XX  
 AC AAC45097;  
 XX  
 DT 18-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 45288.  
 XX  
 KW Hybridisation assay; genetic mapping; gene expression control;  
 KW protein identification; signal transduction pathway;  
 KW metabolic pathway; promoter; termination sequence; ss.

us-09-899-718a-8.rng

Thu Aug 1 08:30:07 2002

```

XX OS Arabidopsis thaliana.
XX XX
XX PN EP1033405-A2.
XX XX
XX PD 06-SEP-2000.
XX XX
XX PF 25-FEB-2000; 2000EP-0301439.
XX XX
XX XX 99US-0121825.
XX XX 25-FEB-1999; 99US-0123180.
XX XX 05-MAR-1999; 99US-0123548.
XX XX 09-MAR-1999; 99US-0123578.
XX XX 23-MAR-1999; 99US-0126264.
XX XX 25-MAR-1999; 99US-0126785.
XX XX 29-MAR-1999; 99US-0127462.
XX XX 01-APR-1999; 99US-0128234.
XX XX 06-APR-1999; 99US-0128714.
XX XX 08-APR-1999; 99US-0129845.
XX XX 16-APR-1999; 99US-0130077.
XX XX 19-APR-1999; 99US-0130449.
XX XX 21-APR-1999; 99US-0130510.
XX XX 23-APR-1999; 99US-0130891.
XX XX 28-APR-1999; 99US-0131449.
XX XX 30-APR-1999; 99US-0132048.
XX XX 30-APR-1999; 99US-0132407.
XX XX 04-MAY-1999; 99US-0132484.
XX XX 05-MAY-1999; 99US-0132485.
XX XX 06-MAY-1999; 99US-0132486.
XX XX 07-MAY-1999; 99US-0132487.
XX XX 07-MAY-1999; 99US-0132863.
XX XX 11-MAY-1999; 99US-0134256.
XX XX 14-MAY-1999; 99US-0134218.
XX XX 14-MAY-1999; 99US-0134219.
XX XX 14-MAY-1999; 99US-0134221.
XX XX 14-MAY-1999; 99US-0134370.
XX XX 14-MAY-1999; 99US-0134768.
XX XX 19-MAY-1999; 99US-0134941.
XX XX 20-MAY-1999; 99US-0135124.
XX XX 21-MAY-1999; 99US-0135353.
XX XX 21-MAY-1999; 99US-0135629.
XX XX 25-MAY-1999; 99US-0136021.
XX XX 27-MAY-1999; 99US-0136392.
XX XX 28-MAY-1999; 99US-0136782.
XX XX 01-JUN-1999; 99US-0137222.
XX XX 03-JUN-1999; 99US-0137528.
XX XX 04-JUN-1999; 99US-0137502.
XX XX 07-JUN-1999; 99US-0137724.
XX XX 08-JUN-1999; 99US-0138094.
XX XX 10-JUN-1999; 99US-0138540.
XX XX 10-JUN-1999; 99US-0138847.
XX XX 14-JUN-1999; 99US-0139119.
XX XX 16-JUN-1999; 99US-0139432.
XX XX 16-JUN-1999; 99US-0139453.
XX XX 17-JUN-1999; 99US-0139492.
XX XX 18-JUN-1999; 99US-0139454.
XX XX 18-JUN-1999; 99US-0139455.
XX XX 18-JUN-1999; 99US-0139456.
XX XX 18-JUN-1999; 99US-0139457.
XX XX 18-JUN-1999; 99US-0139458.
XX XX 18-JUN-1999; 99US-0139459.
XX XX 18-JUN-1999; 99US-0139461.
XX XX 18-JUN-1999; 99US-0139462.
XX XX 18-JUN-1999; 99US-0139463.
XX XX 18-JUN-1999; 99US-0139750.
XX XX 18-JUN-1999; 99US-0139763.
XX XX 21-JUN-1999; 99US-0139817.
XX XX 22-JUN-1999; 99US-0139899.
XX XX 23-JUN-1999; 99US-0140353.
XX XX 23-JUN-1999; 99US-0140354.
XX XX 24-JUN-1999; 99US-0140695.
XX XX 24-JUN-1999; 99US-0140823.
XX XX
XX XX 29-JUN-1999; 99US-0140991.
XX XX 30-JUN-1999; 99US-0141287.
XX XX 01-JUL-1999; 99US-0141842.
XX XX 01-JUL-1999; 99US-0142154.
XX XX 02-JUL-1999; 99US-0142055.
XX XX 06-JUL-1999; 99US-0142390.
XX XX 08-JUL-1999; 99US-0142803.
XX XX 09-JUL-1999; 99US-0142920.
XX XX 12-JUL-1999; 99US-0142977.
XX XX 13-JUL-1999; 99US-0143542.
XX XX 14-JUL-1999; 99US-0143624.
XX XX 15-JUL-1999; 99US-0144005.
XX XX 16-JUL-1999; 99US-0144085.
XX XX 16-JUL-1999; 99US-0144086.
XX XX 19-JUL-1999; 99US-0144325.
XX XX 19-JUL-1999; 99US-0144331.
XX XX 19-JUL-1999; 99US-0144332.
XX XX 19-JUL-1999; 99US-0144333.
XX XX 19-JUL-1999; 99US-0144334.
XX XX 19-JUL-1999; 99US-0144335.
XX XX 20-JUL-1999; 99US-0144352.
XX XX 20-JUL-1999; 99US-0144632.
XX XX 20-JUL-1999; 99US-0144884.
XX XX 21-JUL-1999; 99US-0144814.
XX XX 21-JUL-1999; 99US-0145086.
XX XX 21-JUL-1999; 99US-0145088.
XX XX 21-JUL-1999; 99US-0145085.
XX XX 22-JUL-1999; 99US-0145087.
XX XX 22-JUL-1999; 99US-0145089.
XX XX 22-JUL-1999; 99US-0145192.
XX XX 22-JUL-1999; 99US-0145145.
XX XX 23-JUL-1999; 99US-0145218.
XX XX 23-JUL-1999; 99US-0145224.
XX XX 26-JUL-1999; 99US-0145276.
XX XX 27-JUL-1999; 99US-0145913.
XX XX 27-JUL-1999; 99US-0145918.
XX XX 27-JUL-1999; 99US-0145919.
XX XX 28-JUL-1999; 99US-0145951.
XX XX 02-AUG-1999; 99US-0146386.
XX XX 02-AUG-1999; 99US-0146388.
XX XX 02-AUG-1999; 99US-0146389.
XX XX 03-AUG-1999; 99US-0147038.
XX XX 04-AUG-1999; 99US-0147204.
XX XX 04-AUG-1999; 99US-0147302.
XX XX 05-AUG-1999; 99US-0147192.
XX XX 05-AUG-1999; 99US-0147260.
XX XX 06-AUG-1999; 99US-0147303.
XX XX 06-AUG-1999; 99US-0147416.
XX XX 09-AUG-1999; 99US-0147935.
XX XX 09-AUG-1999; 99US-0148171.
XX XX 10-AUG-1999; 99US-0148319.
XX XX 11-AUG-1999; 99US-0148341.
XX XX 12-AUG-1999; 99US-0148565.
XX XX 13-AUG-1999; 99US-0148684.
XX XX 13-AUG-1999; 99US-0149368.
XX XX 16-AUG-1999; 99US-0149175.
XX XX 17-AUG-1999; 99US-0149426.
XX XX 18-AUG-1999; 99US-0149722.
XX XX 20-AUG-1999; 99US-0149723.
XX XX 20-AUG-1999; 99US-0149929.
XX XX 23-AUG-1999; 99US-0149902.
XX XX 23-AUG-1999; 99US-0149930.
XX XX 25-AUG-1999; 99US-0150566.
XX XX 26-AUG-1999; 99US-0150884.
XX XX 27-AUG-1999; 99US-0151065.
XX XX 27-AUG-1999; 99US-0151066.
XX XX 27-AUG-1999; 99US-0151080.
XX XX 30-AUG-1999; 99US-0151303.
XX XX 31-AUG-1999; 99US-0151438.
XX XX 01-SEP-1999; 99US-0151930.
XX XX 07-SEP-1999; 99US-0152363.
XX XX 10-SEP-1999; 99US-0153070.

```



```

PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154778.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      82.2%; Score 14.8; DB 21; Length 1554;
Best Local Similarity 88.9%; Pred. No. 2.8e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atactctggtcatgttaa 18
   || ||||| |||||
Db 264 atcctctggtcatgttaa 281

RESULT 12
ABLT25967/c
ID ABL25967 standard; DNA; 1926 BP.
XX AC ABL25967;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 29374.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX OS pharmaceutical; gene; ds.
XX PN Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.

```

```

XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
DR New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Claim 1; SEQ ID NO 29374; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI6175) and the encoded proteins
CC (ABBS7737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 1926 BP; 394 A; 531 C; 565 G; 436 T; 0 other;

Query Match      82.2%; Score 14.8; DB 23; Length 1926;
Best Local Similarity 88.9%; Pred. No. 2.8e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atactctggtcatgttaa 18
   || ||||| |||||
Db 1570 ACACGTGTCGTATGTAA 1553

RESULT 13
AAZ30413
ID AAZ30413 standard; DNA; 1935 BP.
XX AC AAZ30413;
XX DT 21-DEC-1999 (first entry)
XX DE Arabidopsis thaliana Mlo fungal resistance gene CIB10295.
XX KW Consensus; resistance; fungus; pathogen; wheat; cell wall; apposition;
XX KW papilla; contact site; callose; carbohydrate; phenol; transgenic plant;
XX KW Mlo; Erysiphe graminis; powdery mildew; ss.
XX OS Arabidopsis thaliana.
XX FH Key Location/Qualifiers
FT CDS 79..1810
FT FT /*tag= a
FT FT /product= "Mlo fungal resistance protein"
XX WO9947552-A2.
XX 23-SEP-1999.
XX 17-MAR-1999; 99WO-EP01779.
XX 17-MAR-1998; 98US-0042763.
XX (NOVS ) NOVARTIS AG.
XX (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX Salmeron JM, Weislo LJ, Strawn LJ, Kramer CM;
PI

```

100

[illegible][illegible]

|    |              |                 |    |                                                                         |                 |
|----|--------------|-----------------|----|-------------------------------------------------------------------------|-----------------|
| PR | 14-AUG-2000; | 2000US-0225759. | PR | 08-NOV-2000;                                                            | 2000US-0246611. |
| PR | 18-AUG-2000; | 2000US-0226279. | PR | 08-NOV-2000;                                                            | 2000US-0246613. |
| PR | 22-AUG-2000; | 2000US-0226681. | PR | 17-NOV-2000;                                                            | 2000US-0249207. |
| PR | 22-AUG-2000; | 2000US-0226868. | PR | 17-NOV-2000;                                                            | 2000US-0249208. |
| PR | 23-AUG-2000; | 2000US-0227182. | PR | 17-NOV-2000;                                                            | 2000US-0249209. |
| PR | 30-AUG-2000; | 2000US-0227009. | PR | 17-NOV-2000;                                                            | 2000US-0249210. |
| PR | 01-SEP-2000; | 2000US-0228924. | PR | 17-NOV-2000;                                                            | 2000US-0249211. |
| PR | 01-SEP-2000; | 2000US-0229287. | PR | 17-NOV-2000;                                                            | 2000US-0249212. |
| PR | 01-SEP-2000; | 2000US-0229343. | PR | 17-NOV-2000;                                                            | 2000US-0249213. |
| PR | 01-SEP-2000; | 2000US-0229344. | PR | 17-NOV-2000;                                                            | 2000US-0249214. |
| PR | 01-SEP-2000; | 2000US-0229345. | PR | 17-NOV-2000;                                                            | 2000US-0249215. |
| PR | 05-SEP-2000; | 2000US-0229345. | PR | 17-NOV-2000;                                                            | 2000US-0249216. |
| PR | 05-SEP-2000; | 2000US-0229509. | PR | 17-NOV-2000;                                                            | 2000US-0249217. |
| PR | 05-SEP-2000; | 2000US-0229513. | PR | 17-NOV-2000;                                                            | 2000US-0249218. |
| PR | 06-SEP-2000; | 2000US-0230437. | PR | 17-NOV-2000;                                                            | 2000US-0249244. |
| PR | 06-SEP-2000; | 2000US-0230438. | PR | 17-NOV-2000;                                                            | 2000US-0249245. |
| PR | 08-SEP-2000; | 2000US-0231242. | PR | 17-NOV-2000;                                                            | 2000US-0249264. |
| PR | 08-SEP-2000; | 2000US-0231243. | PR | 17-NOV-2000;                                                            | 2000US-0249265. |
| PR | 08-SEP-2000; | 2000US-0231244. | PR | 17-NOV-2000;                                                            | 2000US-0249297. |
| PR | 08-SEP-2000; | 2000US-0231413. | PR | 17-NOV-2000;                                                            | 2000US-0249299. |
| PR | 08-SEP-2000; | 2000US-0231414. | PR | 17-NOV-2000;                                                            | 2000US-0249300. |
| PR | 08-SEP-2000; | 2000US-0232080. | PR | 01-DEC-2000;                                                            | 2000US-0250160. |
| PR | 12-SEP-2000; | 2000US-0232081. | PR | 01-DEC-2000;                                                            | 2000US-0250391. |
| PR | 14-SEP-2000; | 2000US-0231968. | PR | 05-DEC-2000;                                                            | 2000US-0251030. |
| PR | 14-SEP-2000; | 2000US-0232397. | PR | 05-DEC-2000;                                                            | 2000US-0251988. |
| PR | 14-SEP-2000; | 2000US-0233398. | PR | 06-DEC-2000;                                                            | 2000US-0251479. |
| PR | 14-SEP-2000; | 2000US-0233399. | PR | 08-DEC-2000;                                                            | 2000US-0251856. |
| PR | 14-SEP-2000; | 2000US-0233400. | PR | 08-DEC-2000;                                                            | 2000US-0251868. |
| PR | 14-SEP-2000; | 2000US-0233401. | PR | 08-DEC-2000;                                                            | 2000US-0251869. |
| PR | 14-SEP-2000; | 2000US-0233063. | PR | 08-DEC-2000;                                                            | 2000US-0251989. |
| PR | 14-SEP-2000; | 2000US-0233064. | PR | 11-DEC-2000;                                                            | 2000US-0251990. |
| PR | 21-SEP-2000; | 2000US-0233065. | PR | 05-JAN-2001;                                                            | 2001US-0254097. |
| PR | 21-SEP-2000; | 2000US-0234223. | XX |                                                                         |                 |
| PR | 25-SEP-2000; | 2000US-0234274. | PA | (HUMA-) HUMAN GENOME SCI INC.                                           |                 |
| PR | 25-SEP-2000; | 2000US-0234957. | XX |                                                                         |                 |
| PR | 26-SEP-2000; | 2000US-0234958. | PI | Rosen CA, Barash SC, Ruben SM;                                          |                 |
| PR | 27-SEP-2000; | 2000US-0235834. | XX | WPI; 2001-483426/52.                                                    |                 |
| PR | 29-SEP-2000; | 2000US-0235836. | XX |                                                                         |                 |
| PR | 29-SEP-2000; | 2000US-0236327. | XX |                                                                         |                 |
| PR | 29-SEP-2000; | 2000US-0236367. | XX |                                                                         |                 |
| PR | 29-SEP-2000; | 2000US-0236368. | XX |                                                                         |                 |
| PR | 29-SEP-2000; | 2000US-0236369. | XX |                                                                         |                 |
| PR | 29-SEP-2000; | 2000US-0236370. | XX |                                                                         |                 |
| PR | 02-OCT-2000; | 2000US-0236802. | PT | Nucleic acids encoding human immune/hematopoietic antigen polypeptides, |                 |
| PR | 02-OCT-2000; | 2000US-0237037. | PT | useful for preventing, diagnosing and/or treating cancers and           |                 |
| PR | 02-OCT-2000; | 2000US-0237038. | XX | metastasis -                                                            |                 |
| PR | 02-OCT-2000; | 2000US-0237039. | XX | Disclosure; SEQ ID NO 22552; 3071pp + Sequence Listing; English.        |                 |
| PR | 02-OCT-2000; | 2000US-0237040. | XX |                                                                         |                 |
| PR | 13-OCT-2000; | 2000US-0239935. | XX |                                                                         |                 |
| PR | 13-OCT-2000; | 2000US-0239937. | XX |                                                                         |                 |
| PR | 20-OCT-2000; | 2000US-0240960. | XX |                                                                         |                 |
| PR | 20-OCT-2000; | 2000US-0241121. | XX |                                                                         |                 |
| PR | 20-OCT-2000; | 2000US-0241785. | XX |                                                                         |                 |
| PR | 20-OCT-2000; | 2000US-0241786. | XX |                                                                         |                 |
| PR | 20-OCT-2000; | 2000US-0241787. | XX |                                                                         |                 |
| PR | 20           |                 |    |                                                                         |                 |

us-09-899-718a-8.rng

Thu Aug 1 08:30:07 2002

Db 474 atacttggtcatgtaa 491  
||||| || |||||||||

Search completed: July 31, 2002, 14:11:28  
Job time: 17385 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:02:11 ; Search time 6034.22 Seconds  
(without alignments)  
62.424 Million cell updates/sec

Title: US-09-899-718A-8

Perfect score: 18

Sequence: 1 atactctggatcgttaa 18

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 1000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_hgt.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sv.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pi.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_hgt\_hum.\*
- 31: em\_hgt\_inv.\*
- 32: em\_hgt\_other.\*
- 33: em\_hgtgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Match | Length | ID         | Description                                                    |
|------------|-------|-------|--------|------------|----------------------------------------------------------------|
| 1          | 18    | 100.0 | 18     | AX349070   | Sequence 8 from Patent WO0202785.                              |
| 2          | 16.4  | 91.1  | 51019  | AX349063   | Sequence 18 bp DNA linear PAT 06-FEB-2002                      |
| 3          | 16.4  | 91.1  | 51019  | AX349070   | Sequence 8 from Patent WO0202785.                              |
| 4          | 16.4  | 91.1  | 55613  | AX349070.1 | GI:18615105                                                    |
| 5          | 16.4  | 91.1  | 73309  |            | synthetic construct.                                           |
| 6          | 16.4  | 91.1  | 81017  |            | synthetic construct.                                           |
| 7          | 16.4  | 91.1  | 105939 |            | artificial sequence.                                           |
| 8          | 16.4  | 91.1  | 110000 |            | 1 (sites)                                                      |
| 9          | 16.4  | 91.1  | 137140 |            | Sprunck S., Kluth, A., Becker, D., Luetticke, S. and Loerz, H. |
| 10         | 16.4  | 91.1  | 138716 |            | Promoters of gene expression in plant caryopses                |
| 11         | 16.4  | 91.1  | 147267 |            | Patent: WO 0202785-A 8 10-JAN-2002;                            |
| 12         | 16.4  | 91.1  | 148418 |            | Aventis CropScience GmbH (DE)                                  |
| 13         | 16.4  | 91.1  | 164235 |            | Location/Qualifiers                                            |
| 14         | 16.4  | 91.1  | 168922 |            | 1. .18                                                         |
| 15         | 16.4  | 91.1  | 176355 |            | /organism="synthetic construct"                                |
| 16         | 16.4  | 91.1  | 182848 |            | /db_xref="taxon:32630"                                         |
| 17         | 16.4  | 91.1  | 184672 |            | /note="Oligonucleotide"                                        |
| 18         | 16.4  | 91.1  | 186848 |            | 5 a 3 c 3 g 7 t                                                |
| 19         | 16.4  | 91.1  | 190160 |            |                                                                |
| 20         | 16.4  | 91.1  | 190376 |            |                                                                |
| 21         | 16.4  | 91.1  | 197069 |            |                                                                |
| 22         | 16.4  | 91.1  | 198540 |            |                                                                |
| 23         | 16.4  | 91.1  | 229041 |            |                                                                |
| 24         | 16    | 88.9  | 19532  |            |                                                                |
| 25         | 16    | 88.9  | 92496  |            |                                                                |
| 26         | 16    | 88.9  | 126247 |            |                                                                |
| 27         | 16    | 88.9  | 148295 |            |                                                                |
| 28         | 16    | 88.9  | 149988 |            |                                                                |
| 29         | 16    | 88.9  | 154618 |            |                                                                |
| 30         | 16    | 88.9  | 164180 |            |                                                                |
| 31         | 16    | 88.9  | 171692 |            |                                                                |
| 32         | 16    | 88.9  | 174472 |            |                                                                |
| 33         | 16    | 88.9  | 196848 |            |                                                                |
| 34         | 15.4  | 85.6  | 1709   |            |                                                                |
| 35         | 15.4  | 85.6  | 1770   |            |                                                                |
| 36         | 15.4  | 85.6  | 1897   |            |                                                                |
| 37         | 15.4  | 85.6  | 1973   |            |                                                                |
| 38         | 15.4  | 85.6  | 2826   |            |                                                                |
| 39         | 15.4  | 85.6  | 3295   |            |                                                                |
| 40         | 15.4  | 85.6  | 41947  |            |                                                                |
| 41         | 15.4  | 85.6  | 43284  |            |                                                                |
| 42         | 15.4  | 85.6  | 50000  |            |                                                                |
| 43         | 15.4  | 85.6  | 53949  |            |                                                                |
| 44         | 15.4  | 85.6  | 65020  |            |                                                                |
| 45         | 15.4  | 85.6  | 67895  |            |                                                                |

|    |      |       |        |    |            |                     |
|----|------|-------|--------|----|------------|---------------------|
| 1  | 18   | 100.0 | 18     | 6  | AX349070   | Sequence            |
| 2  | 18   | 100.0 | 3785   | 5  | AX349063   | Sequence            |
| 3  | 16.4 | 91.1  | 51019  | 5  | AL591420   | Sequence            |
| 4  | 16.4 | 91.1  | 55613  | 9  | AL512443   | Human DNA           |
| 5  | 16.4 | 91.1  | 73309  | 2  | AL606722   | Danio rer           |
| 6  | 16.4 | 91.1  | 81017  | 2  | AC025010   | Homo sapi           |
| 7  | 16.4 | 91.1  | 105939 | 2  | AC094963   | Homo sapi           |
| 8  | 16.4 | 91.1  | 110000 | 2  | AL359456   | Rattus no           |
| 9  | 16.4 | 91.1  | 137140 | 2  | AC091696   | Continuation (2 of  |
| 10 | 16.4 | 91.1  | 138716 | 2  | AL591028   | Felis cat           |
| 11 | 16.4 | 91.1  | 147267 | 2  | AC097356   | AL591028 Homo sapi  |
| 12 | 16.4 | 91.1  | 148418 | 9  | AC020698   | Homo sapi           |
| 13 | 16.4 | 91.1  | 164235 | 30 | AC036175   | AC020698 Homo sapi  |
| 14 | 16.4 | 91.1  | 168922 | 2  | AC108072   | AC036175 Homo sapi  |
| 15 | 16.4 | 91.1  | 176355 | 9  | AL356053   | AC108072 Homo sapi  |
| 16 | 16.4 | 91.1  | 182848 | 2  | AC106149   | AL356053 Human DNA  |
| 17 | 16.4 | 91.1  | 184672 | 2  | AF002508   | AL106149 Rattus no  |
| 18 | 16.4 | 91.1  | 186848 | 2  | AC009272   | AP002508 Homo sapi  |
| 19 | 16.4 | 91.1  | 190160 | 2  | AF001837   | AC009272 Homo sapi  |
| 20 | 16.4 | 91.1  | 190376 | 2  | AC090229   | AF001837 Homo sapi  |
| 21 | 16.4 | 91.1  | 197069 | 2  | AC090370   | AC090229 Homo sapi  |
| 22 | 16.4 | 91.1  | 198540 | 9  | AC019106   | AC090370 Homo sapi  |
| 23 | 16.4 | 91.1  | 229041 | 2  | AC099697   | AC019106 Homo sapi  |
| 24 | 16   | 88.9  | 19532  | 9  | AL596243   | AC099697 Mus muscu  |
| 25 | 16   | 88.9  | 92496  | 9  | AL589848   | AL596243 Human DNA  |
| 26 | 16   | 88.9  | 126247 | 2  | AL591955   | AL589848 Human DNA  |
| 27 | 16   | 88.9  | 148295 | 2  | AL662913   | AC091955 Homo sapi  |
| 28 | 16   | 88.9  | 149988 | 2  | AC011929   | AL662913 Homo sapi  |
| 29 | 16   | 88.9  | 154618 | 2  | AC073403   | AC011929 Homo sapi  |
| 30 | 16   | 88.9  | 164180 | 9  | AC016725   | AC073403 Homo sapi  |
| 31 | 16   | 88.9  | 171692 | 2  | AL662896   | AC016725 Homo sapi  |
| 32 | 16   | 88.9  | 174472 | 2  | AC094034   | AL662896 Homo sapi  |
| 33 | 16   | 88.9  | 196848 | 9  | AC007684   | AC094034 Rattus no  |
| 34 | 15.4 | 85.6  | 1709   | 10 | AF176522   | AC007684 Homo sapi  |
| 35 | 15.4 | 85.6  | 1770   | 5  | CCZP2A     | AF176522 Mus muscu  |
| 36 | 15.4 | 85.6  | 1897   | 5  | CCZP2C     | Z72491 C. carpio mr |
| 37 | 15.4 | 85.6  | 1973   | 10 | AB041571   | Z72493 C. carpio mr |
| 38 | 15.4 | 85.6  | 2826   | 4  | BT250379   | AB041571 Mus muscu  |
| 39 | 15.4 | 85.6  | 3295   | 5  | CCZP2GEN   | AJ250379 Bos tauru  |
| 40 | 15.4 | 85.6  | 41947  | 9  | HSAC000358 | Z72494 C. carpio ZP |
| 41 | 15.4 | 85.6  | 43284  | 9  | AC004201   | AJ250379 Bos tauru  |
| 42 | 15.4 | 85.6  | 50000  | 8  | AY013245   | AC004201 Homo sapi  |
| 43 | 15.4 | 85.6  | 53949  | 2  | AC090724   | AY013245 Oryza sat  |
| 44 | 15.4 | 85.6  | 65020  | 2  | AC102336   | AC090724 Homo sapi  |
| 45 | 15.4 | 85.6  | 67895  | 2  | AF004160   | AC102336 Mus muscu  |

## ALIGNMENTS

|            |            |                                                                |                                   |       |        |                 |                 |
|------------|------------|----------------------------------------------------------------|-----------------------------------|-------|--------|-----------------|-----------------|
| RESULT     | 1          | AX349070                                                       | Sequence 8 from Patent WO0202785. | 18 bp | DNA    | linear          | PAT 06-FEB-2002 |
| LOCUS      | AX349070   | Sequence 8 from Patent WO0202785.                              | 18 bp                             | DNA   | linear | PAT 06-FEB-2002 |                 |
| DEFINITION | AX349070   | Sequence 8 from Patent WO0202785.                              | 18 bp                             | DNA   | linear | PAT 06-FEB-2002 |                 |
| ACCESSION  | AX349070   | Sequence 8 from Patent WO0202785.                              | 18 bp                             | DNA   | linear | PAT 06-FEB-2002 |                 |
| VERSION    | AX349070.1 | GI:18615105                                                    |                                   |       |        |                 |                 |
| KEYWORDS   |            | synthetic construct.                                           |                                   |       |        |                 |                 |
| SOURCE     |            | synthetic construct.                                           |                                   |       |        |                 |                 |
| ORGANISM   |            | synthetic construct.                                           |                                   |       |        |                 |                 |
| REFERENCE  |            | 1 (sites)                                                      |                                   |       |        |                 |                 |
| AUTHORS    |            | Sprunck S., Kluth, A., Becker, D., Luetticke, S. and Loerz, H. |                                   |       |        |                 |                 |
| TITLE      |            | Promoters of gene expression in plant caryopses                |                                   |       |        |                 |                 |
| JOURNAL    |            | Patent: WO 0202785-A 8 10-JAN-2002;                            |                                   |       |        |                 |                 |
| FEATURES   |            | Aventis CropScience GmbH (DE)                                  |                                   |       |        |                 |                 |
| source     |            | Location/Qualifiers                                            |                                   |       |        |                 |                 |
|            |            | 1. .18                                                         |                                   |       |        |                 |                 |
|            |            | /organism="synthetic construct"                                |                                   |       |        |                 |                 |
|            |            | /db_xref="taxon:32630"                                         |                                   |       |        |                 |                 |
|            |            | /note="Oligonucleotide"                                        |                                   |       |        |                 |                 |
| BASE COUNT |            | 5 a 3 c 3 g 7 t                                                |                                   |       |        |                 |                 |
| ORIGIN     |            |                                                                |                                   |       |        |                 |                 |

us-09-899-718a-8.rge

Thu Aug 1 08:30:06 2002

regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30): an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) XX-221H6 is from a Zebrafish PAC library

VECTOR: pCYPAC-6  
IMPORTANT: This sequence is not the entire insert of clone XX-221H6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.  
The true right end of clone XX-221H6 is at 51019 in this sequence. The true right end of clone XX-139E19 is at 2000 in this sequence. This clone was isolated from the BUSM1 library (C. Anemiyia) and provided by C. Anemiyia (Anemiyia lab, Virginia Mason Research Center, Seattle, USA).

FEATURES  
Location/Qualifiers

1..51019  
/organism="Danio rerio"  
/db\_xref="taxon:7955"  
/chromosome="7"  
/clone="XX-221H6"  
/clone\_lib="Anemiyia"  
567..630  
/note="16 copies 4 mer at 75% conserved"

588..631  
/note="22 copies 2 mer ta 86% conserved"

1156..1263  
/note="9 copies 12 mer 86% conserved"

1157..1256  
/note="50 copies 2 mer ta 91% conserved"

1163..1260  
/note="7 copies 14 mer 89% conserved"

1165..1256  
/note="23 copies 4 mer tata 91% conserved"

join(2780..2825,3152..3505,4439..4573,5294..5328,5674..5796)  
/gene="SC:d221H6.1"

join(2780..2825,3152..3505,4439..4573,5294..5328,5674..5796)  
/gene="SC:d221H6.1"

match: match: cDNAs: Em:AF318401  
match: proteins: Tr:CAD12598 Tr:AAK60113

/evidence="not experimental"

/product="SC:d221H6.1 (novel immune-type receptor similar to nitr3r.1)"

/protein\_id="CAD21620.1"

/db\_xref="GI:18369621"

/translation="MSLQDCTFFLLTAYGTCEDFIHQPLVVAELGSSVTLPCPH SDDEFTTISWYKHSACKPLLIASPNQSVTVQAFNNTNRFITITAGSNNLSIL HLEEDFPANYCAKFLNIMMPGEGTILLNEDRNIATSVSPSFPTVPVVCILLIIS VISIVMIVLVIQKSRKKTETQLRSQINQIKDDLNIAALHFSKTKPTTSRSMKT IOETIYSETTVH"

3041..3144  
/note="52 copies 2 mer tt 62% conserved"

4000..4019  
/gene="SC:d221H6.1"

/note="Inconsistency in the number of bases in mononucleotide run between subclones."

4019..4060  
/note="3 copies 14 mer 100% conserved"

complement(4020..4061)  
/note="random repeat. Inconsistency in the number of copies of the repeat element between subclones."

5902..5959  
/note="29 copies 2 mer at 81% conserved"

7894..7919

repeat\_region

misc\_feature

repeat\_region

misc\_feature

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

Query Match 100.0%; Score 18; DB 6; Length 18;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 atactctgggtcatgttaa 18  
|||||  
Db 1 ATACTCTGGTCATGTAA 18

RESULT 2  
AX349063 AX349063 3785 bp DNA linear PAT 06-FEB-2002  
LOCUS  
DEFINITION Sequence 1 from Patent WO0202785.  
ACCESSION AX349063  
VERSION AX349063.1 GI:18615098  
KEYWORDS  
SOURCE bread wheat.  
ORGANISM  
Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.

REFERENCE  
AUTHORS Sprunck,S., Kluth,A., Becker,D., Luetticke,S. and Loerz,H.  
TITLE Promoters of gene expression in plant caryopses  
JOURNAL Patent: WO 0202785-A 1 10-JAN-2002;  
Aventis CropScience GmbH (DE)  
LOCATION/Qualifiers  
1..3785  
/organism="Triticum aestivum"  
/db\_xref="taxon:4565"  
945 a 980 c 899 g 961 t

BASE COUNT  
ORIGIN  
Query Match 100.0%; Score 18; DB 6; Length 3785;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 atactctgggtcatgttaa 18  
|||||  
Db 2999 ATACTCTGGTCATGTAA 3016

RESULT 3  
AL591420 51019 bp DNA linear VRT 23-JAN-2002  
LOCUS  
DEFINITION Zebrafish DNA sequence from clone XX-221H6 on chromosome 7 Contains a novel gene similar to nitr3r.1 (novel immune-type receptor 3r.1), a novel gene similar to nitr3 and part of two novel genes similar to nitr's, complete sequence.  
ACCESSION AL591420.5 GI:16973933  
VERSION AL591420.5  
KEYWORDS HTG; immune-type receptor; nitr3; nitr3r.1.  
SOURCE zebrafish.  
ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 51019)  
Hammond,S.  
Direct Submission  
Submitted (10-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humayy@sanger.ac.uk  
On Nov 17, 2001 this sequence version replaced gi:15722150.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all

COMMENT



Thu Aug 1 08:30:06 2002

us-09-899-718a-8.rge

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr1  
RP11-108P2 is from the library RPCR-11.1 constructed by the group of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm

VECTOR: pBACE3.6  
IMPORTANT: This sequence is not the entire insert of clone RP11-108P2 it may be shorter because we sequence overlapping sections only once, except for a short overlap.  
The true left end of clone RP11-38609 is at 53614 in this sequence.  
The true right end of clone RP3-333A15 is at 2000 in this sequence.

#### FEATURES

source  
Location/Qualifiers  
1..55613  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-108P2"  
/clone\_lib="RPCI-11.1"  
18193 a 10125 c 9972 g 17323 t

#### BASE COUNT

ORIGIN

Query Match 91.1%; Score 16.4; DB 9; Length 55613;  
Best Local Similarity 94.4%; Pred. No. 91;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atactctggtcatgttaa 18

||||| |||||||||

Db 31634 ATACTCTGTCATGTAA 31617

#### RESULT

AL606722/c 73309 bp DNA linear HTG 20-DEC-2001  
LOCUS  
DEFINITION  
Danio rerio chromosome 1 clone XX-27N24, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, in unordered pieces.

AL606722  
AL606722.2 GI:17976610  
HTG: HTGS\_PHASE1: HTGS\_DRAFT; HTGS\_FULLTOP.

zebrafish

Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.

1 (sites)

Burton, J.

Direct Submission  
Submitted (19-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Dec 21, 2001 this sequence version replaced gi:15626247.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: d227N24

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 72055 bases at least Q40

Consensus quality: 72273 bases at least Q30

Consensus quality: 72397 bases at least Q20

Insert size: 72609; sum-of-contigs

Insert size: 88868; 5.6% error; agarose-fp

Quality coverage: 11.34x in Q20 bases; sum-of-contigs Quality

coverage: 9.50x in Q20 bases; agarose-fp

\* NOTE: This is a 'working draft' sequence.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

#### FEATURES

source  
Location/Qualifiers  
1..73309  
/organism="Danio rerio"  
/db\_xref="taxon:7955"  
/chromosome="1"  
/clone="XX-27N24"  
/clone\_lib="Amemiya"  
1..3149  
/note="assembly fragment:01316  
fragment chain:1"  
3250..14161  
/note="assembly fragment:00988  
fragment chain:1"  
14262..25306  
/note="assembly fragment:01324  
fragment chain:1"  
25407..29595  
/note="assembly fragment:01886  
fragment chain:1"  
29696..32897  
/note="assembly fragment:01272  
fragment chain:1"  
32798..52684  
/note="assembly fragment:01855  
fragment chain:1"  
52785..66125  
/note="assembly fragment:01033  
fragment chain:1"  
66226..73309  
/note="assembly fragment:00736"  
BASE COUNT 21551 a 14405 c 13666 g 22985 t 702 others

#### ORIGIN

Query Match 91.1%; Score 16.4; DB 2; Length 73309;  
Best Local Similarity 94.4%; Pred. No. 92;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atactctggtcatgttaa 18

||||| |||||||||

Db 32283 ATACTCTGTCATGTAA 32266

#### RESULT

AC025010 81017 bp DNA linear HTG 13-JUL-2000  
LOCUS  
DEFINITION  
Homo sapiens chromosome 4 clone RP11-85D17 map 4, LOW-PASS SEQUENCE  
SAMPLING.

AC025010

AC025010.1 GI:7145054

KEYWORDS  
HTG; HTGS\_PHASE0.

SOURCE  
human.

ORGANISM  
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 4, clone RP11-85D17

Unpublished

2 (bases 1 to 81017)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bada, F.,

Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,

Campopiano, A., Castle, A., Choquel, Y., Colangelo, M., Collins, S.,

Collamore, A., Cooke, P., DeArrelano, K., Dewar, K., Diaz, J.S.,

Dodge, S., Domino, M., Doyle, M., Ferrel, P., FitzHugh, W., Gage, D.,

Gallagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,

Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,



Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
 Meldrum, J., Meneus, L., Minova, I., Miranda, C., Mienga, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE  
 JOURNAL  
 COMMENT

Submitted (03-MAR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Center project name: L7756

Center clone name: 85\_D\_17

\* NOTE: This record contains 94 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1  
 709: contig of 709 bp in length  
 710 809: gap of 100 bp  
 810 1567: contig of 758 bp in length  
 1568 1667: gap of 100 bp  
 1668 2433: contig of 766 bp in length  
 2434 2533: gap of 100 bp  
 2534 3310: contig of 777 bp in length  
 3311 3410: gap of 100 bp  
 3411 4176: contig of 766 bp in length  
 4177 4276: gap of 100 bp  
 4277 5036: contig of 760 bp in length  
 5037 5136: gap of 100 bp  
 5137 5899: contig of 763 bp in length  
 5900 5999: gap of 100 bp  
 6000 6764: contig of 765 bp in length  
 6765 6864: gap of 100 bp  
 6865 7612: contig of 748 bp in length  
 7613 7712: gap of 100 bp  
 7713 8467: contig of 755 bp in length  
 8468 8567: gap of 100 bp  
 8568 9320: contig of 753 bp in length  
 9321 9420: gap of 100 bp  
 9421 10187: contig of 767 bp in length  
 10188 10287: gap of 100 bp  
 10288 11042: contig of 755 bp in length  
 11043 11142: gap of 100 bp  
 11143 11910: contig of 768 bp in length  
 11911 12010: gap of 100 bp  
 12011 12757: contig of 747 bp in length  
 12758 12857: gap of 100 bp  
 12858 13619: contig of 762 bp in length  
 13620 13719: gap of 100 bp  
 13720 14493: contig of 774 bp in length  
 14494 14593: gap of 100 bp  
 14594 15348: contig of 755 bp in length

15349 15448: gap of 100 bp  
 15449 16206: contig of 758 bp in length  
 16207 16306: gap of 100 bp  
 16307 17054: contig of 748 bp in length  
 17055 17154: gap of 100 bp  
 17155 17900: contig of 746 bp in length  
 17901 18000: gap of 100 bp  
 18001 18767: contig of 767 bp in length  
 18768 18867: gap of 100 bp  
 18868 19642: contig of 775 bp in length  
 19643 19742: gap of 100 bp  
 19743 20493: contig of 751 bp in length  
 20494 20593: gap of 100 bp  
 20594 21337: contig of 744 bp in length  
 21338 21437: gap of 100 bp  
 21438 22202: contig of 765 bp in length  
 22203 22302: gap of 100 bp  
 22303 23072: contig of 770 bp in length  
 23073 23172: gap of 100 bp  
 23173 23941: contig of 769 bp in length  
 23942 24041: gap of 100 bp  
 24042 24799: contig of 758 bp in length  
 24800 24899: gap of 100 bp  
 24900 25667: contig of 768 bp in length  
 25668 25767: gap of 100 bp  
 25768 26515: contig of 748 bp in length  
 26516 26615: gap of 100 bp  
 26616 27375: contig of 760 bp in length  
 27376 27475: gap of 100 bp  
 27476 28238: contig of 763 bp in length  
 28239 28338: gap of 100 bp  
 28339 29104: contig of 766 bp in length  
 29105 29204: gap of 100 bp  
 29205 29969: contig of 765 bp in length  
 29970 30069: gap of 100 bp  
 30070 30837: contig of 768 bp in length  
 30838 30937: gap of 100 bp  
 30939 31710: contig of 773 bp in length  
 31711 31810: gap of 100 bp  
 31811 32576: contig of 766 bp in length  
 32577 32676: gap of 100 bp  
 32677 33461: contig of 785 bp in length  
 33462 33561: gap of 100 bp  
 33562 34337: contig of 776 bp in length  
 34338 34437: gap of 100 bp  
 34438 35216: contig of 779 bp in length  
 35217 35316: gap of 100 bp  
 35317 36082: contig of 766 bp in length  
 36083 36182: gap of 100 bp  
 36183 36951: contig of 769 bp in length  
 36952 37051: gap of 100 bp  
 37052 37815: contig of 764 bp in length  
 37816 37915: gap of 100 bp  
 37916 38676: contig of 761 bp in length  
 38677 38776: gap of 100 bp  
 38777 39563: contig of 787 bp in length  
 39564 39663: gap of 100 bp  
 39664 40445: contig of 782 bp in length  
 40446 40545: gap of 100 bp  
 40546 41316: contig of 771 bp in length  
 41317 41416: gap of 100 bp  
 41417 42181: contig of 765 bp in length  
 42182 42281: gap of 100 bp  
 42282 43030: contig of 749 bp in length  
 43031 43130: gap of 100 bp  
 43131 43913: contig of 783 bp in length  
 43914 44013: gap of 100 bp  
 44014 44770: contig of 757 bp in length  
 44771 44870: gap of 100 bp  
 44871 45631: contig of 761 bp in length  
 45632 45731: gap of 100 bp  
 45732 46487: contig of 756 bp in length  
 46488 46587: gap of 100 bp

Thu Aug 1 08:30:06 2002

\* 46588 47340: contig of 753 bp in length  
 \* 47341 47440: gap of 100 bp  
 \* 47441 48171: contig of 731 bp in length  
 \* 48172 48271: gap of 100 bp  
 \* 48272 49028: contig of 757 bp in length  
 \* 49029 49128: gap of 100 bp  
 \* 49129 49883: contig of 755 bp in length  
 \* 49884 49983: gap of 100 bp  
 \* 49984 50748: contig of 765 bp in length  
 \* 50749 50848: gap of 100 bp  
 \* 50849 51619: contig of 771 bp in length  
 \* 51620 51719: gap of 100 bp  
 \* 51720 52473: contig of 754 bp in length  
 \* 52474 52573: gap of 100 bp  
 \* 52574 53327: contig of 754 bp in length  
 \* 53328 53427: gap of 100 bp  
 \* 53428 54198: contig of 771 bp in length  
 \* 54199 54298: gap of 100 bp  
 \* 54299 55054: contig of 756 bp in length  
 \* 55055 55154: gap of 100 bp  
 \* 55155 55921: contig of 767 bp in length  
 \* 55922 56021: gap of 100 bp  
 \* 56022 56781: contig of 760 bp in length  
 \* 56782 56881: gap of 100 bp  
 \* 56882 57638: contig of 757 bp in length  
 \* 57639 57738: gap of 100 bp  
 \* 57739 58492: contig of 754 bp in length  
 \* 58493 58592: gap of 100 bp  
 \* 58593 59364: contig of 772 bp in length  
 \* 59365 59464: gap of 100 bp

Query Match 91.1%; Score 16.4; DB 2; Length 81017;

Best Local Similarity 94.4%; Pred. NO. 92;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atactctggctcgtgtaa 18

Db 12578 AGACTCGTCATGTTAA 12595

RESULT 7  
 AC094963/c  
 LOCUS  
 DEFINITION Rattus norvegicus clone CH230-6L20, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 AC094963 105939 bp DNA linear HTG 20-DEC-2001  
 56 unordered pieces.

AC094963 GI:17941764

HTG: HTGS PHASE1.

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 105939)

Alsbrooks, S.L., Amarantunga, H.C., Are, J.R., Banks, T., Barbaria, J.,  
 Benton, J., Bivage, K., Blankenburg, K., Bonnin, D., Bouck, J.,  
 Bowler, S., Brieve, M., Brown, E., Brown, M., Bryant, N.P., Buha, C.,  
 Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, F.F.,  
 Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,  
 Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,  
 Coyle, M.D., Dathorne, S.K., David, R., Davila, M.L., Davis, C.,  
 Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,  
 Denn, A.L., Ding, J., Dinh, H.H., Douthwaite, K.J., Draper, H.,  
 Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,  
 Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,  
 Foster, P., Franco, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,  
 Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,  
 Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,  
 Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C.,  
 Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,  
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,  
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,

Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,  
 Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,  
 Loulsegue, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,  
 Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M.,  
 Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,  
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
 Nguyen, A., Nguyen, N., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,  
 Ogih, M., Okwuonu, G., Oragunye, N., Pickens, R., Primus, E., Pu, L.,  
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.,  
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojokan, I., Rolfe, M.,  
 Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshkari, N.,  
 Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,  
 Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,  
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,  
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R.,  
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
 Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S.,  
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G. and Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 105939)

Worley, K.C.

Direct Submission

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Dec 20, 2001 this sequence version replaced gi:17062175.

## COMMENT

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GBU

Center clone name: CH230-6L20

----- Summary Statistics

Assembly program: Phrap; version 0.990329First call to

findPhrapList

Consensus quality: 82923 bases at least Q40

Consensus quality: 92369 bases at least Q30

Consensus quality: 99005 bases at least Q20

Estimated insert size: 70929; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 0.8x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

consists of 56 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

4041: contig of 4041 bp in length

4141: gap of unknown length

8685: contig of 4544 bp in length

8686: gap of unknown length

12107: contig of 3322 bp in length

12108: gap of unknown length

15728: contig of 3521 bp in length

15828: gap of unknown length

20121: contig of 4293 bp in length

20222: gap of unknown length

22826: contig of 2605 bp in length

22927: gap of unknown length

25071: contig of 2145 bp in length

25171: gap of unknown length

28657: contig of 3486 bp in length

28757: gap of unknown length

\* 28758 31090: contig of 2333 bp in length  
\* 31091 31190: gap of unknown length  
\* 31191 33330: contig of 2140 bp in length  
\* 33331 33430: gap of unknown length  
\* 33431 35098: contig of 1668 bp in length  
\* 35099 35198: gap of unknown length  
\* 35199 37556: contig of 2358 bp in length  
\* 37557 37656: gap of unknown length  
\* 37657 38712: contig of 1056 bp in length  
\* 38713 38812: gap of unknown length  
\* 38813 40216: contig of 1404 bp in length  
\* 40217 40316: gap of unknown length  
\* 40317 42189: contig of 1873 bp in length  
\* 42190 42289: gap of unknown length  
\* 42290 44826: contig of 2537 bp in length  
\* 44827 44926: gap of unknown length  
\* 44927 46654: contig of 1728 bp in length  
\* 46655 46754: gap of unknown length  
\* 46755 48630: contig of 1876 bp in length  
\* 48631 48730: gap of unknown length  
\* 48731 50244: contig of 1514 bp in length  
\* 50245 50344: gap of unknown length  
\* 50345 52054: contig of 1710 bp in length  
\* 52055 52154: gap of unknown length  
\* 52155 53878: contig of 1724 bp in length  
\* 53879 53978: gap of unknown length  
\* 53979 55533: contig of 1555 bp in length  
\* 55534 55633: gap of unknown length  
\* 55634 57398: contig of 1765 bp in length  
\* 57399 57498: gap of unknown length  
\* 57499 59018: contig of 1520 bp in length  
\* 59019 59118: gap of unknown length  
\* 59119 60951: contig of 1833 bp in length  
\* 60952 61051: gap of unknown length  
\* 61052 63791: contig of 2740 bp in length  
\* 63792 63891: gap of unknown length  
\* 63892 64936: contig of 1045 bp in length  
\* 64937 65036: gap of unknown length  
\* 65037 65337: contig of 1497 bp in length  
\* 65338 66333: gap of unknown length  
\* 66334 68166: contig of 1533 bp in length  
\* 68167 68266: gap of unknown length  
\* 68267 69395: contig of 1129 bp in length  
\* 69396 69495: gap of unknown length  
\* 69496 70721: contig of 1226 bp in length  
\* 70722 70821: gap of unknown length  
\* 70822 72079: contig of 1258 bp in length  
\* 72080 72179: gap of unknown length  
\* 72180 73300: contig of 1121 bp in length  
\* 73301 73400: gap of unknown length  
\* 73401 74409: contig of 1009 bp in length  
\* 74410 74509: gap of unknown length  
\* 74510 76335: contig of 1826 bp in length  
\* 76336 76435: gap of unknown length  
\* 76436 77982: contig of 1547 bp in length  
\* 77983 78082: gap of unknown length  
\* 78083 79545: contig of 1463 bp in length  
\* 79546 79645: gap of unknown length  
\* 79646 81116: contig of 1471 bp in length  
\* 81117 82544: contig of 1328 bp in length  
\* 82545 82644: gap of unknown length  
\* 82645 83693: contig of 1049 bp in length  
\* 83694 83793: gap of unknown length  
\* 83794 85175: contig of 1382 bp in length  
\* 85176 85275: gap of unknown length  
\* 85276 86349: contig of 1074 bp in length  
\* 86350 86450: gap of unknown length  
\* 86450 87888: contig of 1439 bp in length  
\* 87889 89628: gap of unknown length  
\* 89629 89728: gap of unknown length  
\* 89729 91699: contig of 1971 bp in length

\* 91700 91799: gap of unknown length  
\* 91800 92828: contig of 1029 bp in length  
\* 92829 92928: gap of unknown length  
\* 92929 94150: contig of 1222 bp in length  
\* 94151 94250: gap of unknown length  
\* 94251 95381: contig of 1131 bp in length  
\* 95382 95481: gap of unknown length  
\* 95482 96491: contig of 1010 bp in length  
\* 96492 96591: gap of unknown length  
\* 96592 97820: contig of 1228 bp in length  
\* 97821 97919: gap of unknown length  
\* 97920 99613: contig of 1594 bp in length  
\* 99614 100697: contig of 1084 bp in length  
\* 100698 100798: gap of unknown length  
\* 100799 102235: contig of 1438 bp in length  
\* 102236 102335: gap of unknown length  
\* 102336 103409: contig of 1074 bp in length  
\* 103410 103509: gap of unknown length  
\* 103510 104544: contig of 1035 bp in length  
\* 104545 104644: gap of unknown length  
\* 104645 105939: contig of 1295 bp in length.

## FEATURES

Location/Qualifiers

Query Match 91.1%; Score 16.4; DB 2; Length 105939;  
Best Local Similarity 94.4%; Pred. No. 94;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctgtcatgttaa 18

Db 97179 ATACTCTGTCATGTAA 97162

## RESULT 8

AL359456\_1  
WPCOMMENT

Sequence split into 6 fragments LOCUS AL359456 Accession AL359456

| Fragment Name | Begin  | End    |
|---------------|--------|--------|
| AL359456_0    | 1      | 110000 |
| AL359456_1    | 100001 | 210000 |
| AL359456_2    | 200001 | 310000 |
| AL359456_3    | 300001 | 410000 |
| AL359456_4    | 400001 | 510000 |
| AL359456_5    | 500001 | 593964 |

Continuation (2 of 6) of AL359456 from base 100001 (AL359456 Homo sapiens chromosome

## Query Match

Best Local Similarity 91.1%; Score 16.4; DB 2; Length 110000;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atactctgtcatgttaa 18

Db 13235 ATACTCTGTCATGTAA 13252

## RESULT 9

AC091696/c

LOCUS

DEFINITION AC091696 137140 bp DNA linear HTG 19-MAY-2001  
Felis catus clone RP86-117J4, WORKING DRAFT SEQUENCE, 8 unordered

AC091696

pieces

AC091696

VERSION AC091696.1 GI:14150418

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE cat.

ORGANISM Felis catus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.

1 (bases 1 to 137140)

Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,

Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S.,

Guan, X., Gupta, J., Ho, S.-L., Idol, J.R., Karlins, E., Lee-Lin, S.-Q.,

Legaspi, R., Lim, M., Maduro, Q.L., Maduro, V.B., Masiello, C.,

us-09-899-718a-8.1rge

Thu Aug 1 08:30:06 2002

Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A.,  
 Shevchenko, Y., Snyder, B., Stantrick, S., Thomas, J.W., Thomas, P.J.,  
 Tjongson, E.E., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A.,  
 Wetherby, K.D., Zhang, L.-H. and Green, E.D.  
 NISC Comparative Sequencing Initiative  
 Unpublished  
 2 (bases 1 to 137140)  
 Green, E.D.  
 Direct Submission  
 Submitted (19-MAY-2001) NIH Intramural Sequencing Center, 8717  
 Government Circle, Gaithersburg, MD 20877, USA  
 ----- Genome Center  
 Center: NIH Intramural Sequencing Center  
 Center code: NISC  
 Web site: <http://www.nisc.nih.gov>  
 Contact: [nisc\\_mouse@nhgri.nih.gov](mailto:nisc_mouse@nhgri.nih.gov)  
 ----- Project Information  
 ----- Project name: awg  
 Center project name: 117304  
 Center clone name: 117304  
 ----- Summary Statistics  
 Sequencing vector: plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 133760 bases at least Q40  
 Consensus quality: 134423 bases at least Q30  
 Consensus quality: 134718 bases at least Q20  
 Insert size: 133000; agarose-fp  
 Quality coverage: 11.63x in Q20 bases; agarose-fp  
 Quality coverage: 11.33x in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 2316: contig of 2316 bp in length  
 \* 2317 2416: gap of unknown length  
 \* 2417 13393: contig of 10977 bp in length  
 \* 13394 13493: gap of unknown length  
 \* 13494 20080: contig of 6587 bp in length  
 \* 20081 20180: gap of unknown length  
 \* 20181 31115: contig of 10935 bp in length  
 \* 31116 31215: gap of unknown length  
 \* 31216 47257: contig of 16042 bp in length  
 \* 47258 47357: gap of unknown length  
 \* 47358 72019: contig of 24662 bp in length  
 \* 72020 72119: gap of unknown length  
 \* 72120 102668: contig of 30549 bp in length  
 \* 102669 102769: gap of unknown length  
 \* 102769 137140: contig of 34372 bp in length.

## FEATURES

Source  
 1. 137140  
 /organism="Felis catus"  
 /db\_xref="taxon:9685"  
 /clone="RP86-117J4"  
 /clone\_lib="RP86"  
 misc\_feature  
 1. 2316  
 /note="assembly\_fragment"  
 misc\_feature  
 2417. 13393  
 /note="assembly\_fragment"  
 misc\_feature  
 13494. 20080  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:left  
 misc\_feature  
 20181. 31115  
 /note="assembly\_fragment"  
 misc\_feature  
 31216. 47257  
 /note="assembly\_fragment"  
 clone\_end:T7

vector\_side:right  
 47358. 72019  
 /note="assembly\_fragment"  
 misc\_feature  
 72120. 102668  
 /note="assembly\_fragment"  
 misc\_feature  
 102769. 137140  
 /note="assembly\_fragment"  
 BASE COUNT 37260 a 30755 c 29908 g 38508 t 709 others  
 ORIGIN

Query Match 91.1%; Score 16.4; DB 2; Length 137140;  
 Best Local Similarity 94.4%; Pred. No. 95;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggctcatgttaa 18  
 |||||  
 Db 65046 ATACTCTGGTCATGTGA 65029

RESULT 10  
 AL591028/c 138716 bp DNA linear HTG 19-DEC-2001  
 LOCUS Homo sapiens chromosome 10 clone RP13-143F22, \*\*\* SEQUENCING IN  
 DEFINITION PROGRESS \*\*\*, 11 unordered pieces.  
 AL591028  
 ACCESSION AL591028.3 GI:17973956  
 VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS human.  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (Sites)  
 Burton, J.  
 Direct Submission  
 Submitted (18-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
[humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk) Clone requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
 On Dec 20, 2001 this sequence version replaced gi:14133148.  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
 ----- Project Information  
 Center project name: b143F22  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: plasmid; 108752; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Consensus quality: 134354 bases at least Q40  
 Consensus quality: 136114 bases at least Q30  
 Consensus quality: 137107 bases at least Q20  
 Insert size: 137716; sum-of-contigs  
 Insert size: 151380; 17.0% error; agarose-fp  
 Quality coverage: 5.19x in Q20 bases; sum-of-contigs Quality  
 coverage: 4.77x in Q20 bases; agarose-fp

----- NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 11 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 2424: contig of 2424 bp in length  
 \* 2425 2524: gap of 100 bp  
 \* 2525 24808: contig of 22284 bp in length  
 \* 24809 24908: gap of 100 bp  
 \* 24909 35362: contig of 10454 bp in length  
 \* 35363 35462: gap of 100 bp  
 \* 35463 43765: contig of 8303 bp in length

```

* 43766 43865: gap of 100 bp
* 43866 47699: contig of 3834 bp in length
* 47700 47799: gap of 100 bp
* 47800 78351: contig of 30552 bp in length
* 78352 78451: gap of 100 bp
* 78452 81539: contig of 3088 bp in length
* 81540 81639: gap of 100 bp
* 81640 88953: contig of 7314 bp in length
* 88954 89053: gap of 100 bp
* 89054 95324: contig of 6271 bp in length
* 95325 95424: gap of 100 bp
* 95425 98902: contig of 3478 bp in length
* 98903 99002: gap of 100 bp
* 99003 138716: contig of 39714 bp in length.
FEATURES
    source
        1. .138716
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="10"
            /clone="Rp13-143F22"
            /clone_lib="RPCI-13.1"
            1. .2424
                /note="assembly_fragment:01914
                fragment_chain:1"
                2525. .24808
                    /note="assembly_fragment:02177
                    fragment_chain:1"
                    24909. .35362
                        /note="assembly_fragment:01947
                        fragment_chain:1"
                        35463. .43765
                            /note="assembly_fragment:00801
                            fragment_chain:1"
                            43866. .47699
                                /note="assembly_fragment:01909
                                fragment_chain:1"
                                47800. .78351
                                    /note="assembly_fragment:00910
                                    fragment_chain:2"
                                    78452. .81539
                                        /note="assembly_fragment:02178
                                        fragment_chain:2"
                                        81640. .88953
                                            /note="assembly_fragment:00075
                                            fragment_chain:2"
                                            89054. .95324
                                                /note="assembly_fragment:01879
                                                fragment_chain:2"
                                                95425. .98902
                                                    /note="assembly_fragment:01238"
                                                    99003. .138716
                                                        /note="assembly_fragment:01608"
                                                        /note="assembly_fragment:01608"
BASE COUNT 45713 a 26430 c 25963 g 39610 t 1000 others
ORIGIN
Query Match 91.1%; Score 16.4; DB 2; Length 138716;
Best Local Similarity 94.4%; Pred. No. 95;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggtcatgttaa 18
||||| |||||||
Db 55073 ATACACTGGTCATGTAA 55056

RESULT 11
AC097356/c
LOCUS
DEFINITION Homo sapiens chromosome 3 clone RP11-107K23, WORKING DRAFT
SEQUENCE, 3 unordered pieces.
ACCESSION AC097356 AC024017
VERSION AC097356.1 GI:16152306
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.

```

## SOURCE

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 147267)

Kaul, R.K., Olson, M.V., Raymond, C., Clendenning, J. and Haugen, E.D. Direct Submission

Unpublished

2 (bases 1 to 147267)

Kaul, R.K., Olson, M.V., Raymond, C., Clendenning, J. and Haugen, E.D. Direct Submission

Submitted (16-OCT-2001)

Box 352145, Seattle, WA 98195, USA

On Oct 16, 2001 this sequence version replaced gi:7121074.

----- Genome Center

Center: University of Washington Genome Center

Center Code: UWGC

Web site: http://www.genome.washington.edu

Contact: uwgchgs@u.washington.edu

Drafting Center: BCM

----- Project Information

Center project name: chr-3

Center clone name: RP11-107K23 (bc0228)

----- Summary Statistics

Sequencing vector: unknown; 65% of reads

Chemistry: Dye-terminator ET; 66% of reads

Chemistry: Dye-terminator Big Dye; 34% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 145614 bases at least Q40

Consensus quality: 146445 bases at least Q30

Consensus quality: 146882 bases at least Q20

Insert size: 147067; sum-of-contigs

Quality coverage: 9.4x in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 8475: contig of 8475 bp in length

\* 8476 8575: gap of unknown length

\* 8576 47158: contig of 38583 bp in length

\* 47159 47258: gap of unknown length

\* 47259 147267: contig of 100009 bp in length.

FEATURES

## source

1. .147267

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="3"

/clone="RP11-107K23"

/clone\_lib="RPCI human BAC library 11"

1. .8475

misc\_feature

/note="assembly\_name:Contig12"

8576. 47158

misc\_feature

/note="assembly\_name:Contig13"

47259. .147267

misc\_feature

/note="assembly\_name:Contig14"

BASE COUNT 45465 a 29529 c 28742 g 43321 t 210 others

## ORIGIN

Query Match 91.1%; Score 16.4; DB 2; Length 147267;

Best Local Similarity 94.4%; Pred. No. 96;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggtcatgttaa 18

||||| |||||||

Db 49216 ATACTCTGTCATGTAA 49199

Thu Aug 1 08:30:06 2002

us-09-899-718a-8.rge

## RESULT 12

AC020698 148418 bp DNA linear PRI 09-MAY-2001  
 LOCUS Homo sapiens BAC clone RP11-45F23 from 5, complete sequence.  
 DEFINITION AC020698  
 ACCESSION AC020698  
 VERSION AC020698.4 GI:11120934  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 148418)  
 AUTHORS Sulston, J.E. and Waterston, R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99063792  
 REFERENCE 2 (bases 1 to 148418)  
 AUTHORS Nguyen, C., Drone, K., Hawkins, M. and Ureta, M.  
 TITLE The sequence of Homo sapiens BAC clone RP11-45F23  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 148418)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 4 (bases 1 to 148418)  
 AUTHORS Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-NOV-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 REFERENCE 5 (bases 1 to 148418)  
 AUTHORS Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 COMMENT On Nov 8, 2000 this sequence version replaced gi:7630812.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)  
 ----- Summary Statistics  
 -----  
 Center project name: H\_NH0045F23  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPC1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenio, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong

and coworkers at the Roswell Park Cancer Institute  
 (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-462G22. Actual start of this clone is at base position 1 of RP11-45F23; actual end is at base position 148418 of RP11-45F23.

Location/Qualifiers

## FEATURES

source

1. .148418

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="2"

/map="5"

/clone="RP11-45F23"

/clone\_lib="RPC1-11"

1. .434

/rpt\_family="L1"

507. .807

/rpt\_family="L1"

807. .2585

/rpt\_family="L1"

2616. .3020

/rpt\_family="L1"

3062. .3250

/rpt\_family="L1"

3251. .3596

/rpt\_family="MER2\_type"

3597. .3843

/rpt\_family="L1"

9001. .9137

/rpt\_family="Alu"

11353. .11406

/rpt\_family="ERVL"

17191. .17869

/rpt\_family="L1"

19869. .20177

/rpt\_family="Alu"

21287. .22345

/rpt\_family="L1"

22727. .22937

/rpt\_family="MER2\_type"

22933. .23133

/rpt\_family="MER2\_type"

23355. .23442

/rpt\_family="MIR"

23485. .23625

/rpt\_family="MER2\_type"

24107. .24872

/rpt\_family="L1"

24873. .25320

/rpt\_family="ERV1"

25372. .25522

/note="similar to EST BE087288 (NID:98477683)"

28473. .28614

/rpt\_family="ERV1"

28622. .28736

/rpt\_family="ERV1"

28790. .28843

/rpt\_family="ERV1"

28837. .29212

/rpt\_family="ERV1"

29215. .29821

/rpt\_family="ERV1"

29823. .29972

/rpt\_family="ERV1"

29974. .30286

/rpt\_family="ERV1"

30287. .30737

/rpt\_family="ERV1"

30745. .31216

/rpt\_family="L1"

31227. .31275

/rpt\_family="MIR"

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region





Thu Aug 1 08:30:06 2002

\* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1  
 1026: contig of 1026 bp in length  
 1027 1126: gap of 100 bp  
 1127 1306: contig of 180 bp in length  
 1307 1406: gap of 100 bp  
 1407 2680: contig of 1274 bp in length  
 2681 2780: gap of 100 bp  
 2781 4606: contig of 1826 bp in length  
 4607 4706: gap of 100 bp  
 4707 7010: contig of 2304 bp in length  
 7011 7110: gap of 100 bp  
 7111 10959: contig of 3849 bp in length  
 10960 11059: gap of 100 bp  
 11060 15310: contig of 4251 bp in length  
 15311 15410: gap of 100 bp  
 15411 20672: contig of 5262 bp in length  
 20673 20772: gap of 100 bp  
 20773 26614: contig of 5842 bp in length  
 26615 26714: gap of 100 bp  
 26715 32318: contig of 5604 bp in length  
 32319 32418: gap of 100 bp  
 32419 39439: contig of 7021 bp in length  
 39440 39539: gap of 100 bp  
 47838 47937: gap of 100 bp  
 47938 58126: contig of 10189 bp in length  
 58127 58226: gap of 100 bp  
 58227 66313: contig of 8087 bp in length  
 66314 66413: gap of 100 bp  
 66414 78356: contig of 11943 bp in length  
 78357 78456: gap of 100 bp  
 78457 90502: contig of 12046 bp in length  
 90503 90602: gap of 100 bp  
 90603 110197: contig of 19595 bp in length  
 110198 110297: gap of 100 bp  
 110298 132734: contig of 22437 bp in length  
 132735 132834: gap of 100 bp  
 132835 164235: contig of 31401 bp in length.

Key Location/Qualifiers

source 1..164235  
 /chromosome="2"  
 /db\_xref="taxon:9606"  
 /organism="Homo sapiens"  
 /map="2"  
 /clone="RP11-704A16"  
 /clone.lib="RPC1-11 Human Male BAC"

misc\_feature 1..1026  
 /note="assembly\_fragment"  
 misc\_feature 1127..1306  
 /note="assembly\_fragment clone\_end:T7 vector\_side:right"  
 misc\_feature 1407..2680  
 /note="assembly\_fragment"  
 misc\_feature 2781..4606  
 /note="assembly\_fragment"  
 misc\_feature 4707..7010  
 /note="assembly\_fragment"  
 misc\_feature 7111..10959  
 /note="assembly\_fragment"  
 misc\_feature 11060..15310  
 /note="assembly\_fragment"  
 misc\_feature 15411..20672  
 /note="assembly\_fragment"  
 misc\_feature 20773..26614  
 /note="assembly\_fragment"  
 misc\_feature 26715..32318  
 /note="assembly\_fragment"  
 misc\_feature 32419..39439

FT misc\_feature /note="assembly\_fragment"  
 FT 39540..47837  
 FT /note="assembly\_fragment"  
 FT 47938..58126  
 FT /note="assembly\_fragment"  
 FT 58227..66313  
 FT /note="assembly\_fragment clone\_end:SP6 vector\_side:left"  
 FT 66414..78356  
 FT /note="assembly\_fragment"  
 FT 78457..90502  
 FT /note="assembly\_fragment"  
 FT 90603..110197  
 FT /note="assembly\_fragment"  
 FT 110298..132734  
 FT /note="assembly\_fragment"  
 FT 132835..164235  
 FT /note="assembly\_fragment"  
 XX Sequence 164235 BP; 52554 A; 29032 C; 29120 G; 51728 T; 1801 other;  
 SQ

Query Match 91.1%; Score 16.4; DB 30; Length 164235;  
 Best Local Similarity 94.4%; Pred. No. 96;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 atactctgtcatgttaa 18  
 ||| ||||| ||||| |||||  
 DB 95389 ATAATCTGTCATGTAA 95372

RESULT 14  
 AC108072/c  
 LOCUS Homo sapiens chromosome 2 clone RP11-704A16, WORKING DRAFT  
 DEFINITION  
 SEQUENCE, 2 unordered pieces.  
 AC108072 AC036175  
 AC108072.2 GI:18376983  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Waterston,R.H.  
 The sequence of Homo sapiens clone  
 Unpublished  
 2 (bases 1 to 166922)  
 Waterston,R.H.  
 Direct Submission  
 Submitted (24-JAN-2002) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 On Jan 26, 2002 this sequence version replaced gi:18308946.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 AUTHORS  
 TITLE  
 JOURNAL

COMMENT

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 Contact: submissions@wustl.edu  
 ----- Project Information -----  
 Center project name: H\_NH0704A16  
 Drafting center: WIBR  
 ----- Summary Statistics -----  
 Sequencing vector: M13; 43%  
 Chemistry: Dye-primer ET; 0% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 166298 bases at least Q40  
 Consensus quality: 166490 bases at least Q50  
 Consensus quality: 166729 bases at least Q20  
 Insert size: 168000; agarose-fp



Insert size: 166822; sum-of-contigs  
 Quality coverage: 11.20 in Q20 bases; agarose-gp  
 Quality coverage: 11.28 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1113: contig of 1113 bp in length  
 1114 1213: gap of unknown length  
 1214 166922: contig of 165709 bp in length.

## FEATURES

source

1. 166922  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /clone="RP11-704A16"  
 1. 1113  
 /note="assembly\_name:Contig4"  
 1214. 166922  
 /note="assembly\_name:Contig8  
 clone\_end:SP6  
 vector\_side:right"

misc\_feature

misc\_feature

BASE COUNT 54810 a 30046 c 29856 g 52110 t 100 others  
 ORIGIN

Query Match 91.1%; Score 16.4; DB 2; Length 166922;  
 Best Local Similarity 94.4%; Pred. No. 96;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atactctggctcatgttaa 18

|||||

Db 100989 ATACTCTGGTCATGTAA 100972

## RESULT 15

AL356053

LOCUS

DEFINITION Human DNA sequence from clone RP11-453G10 on chromosome 10,  
 complete sequence.

ACCESSION AL356053

VERSION AL356053.14

KEYWORDS GI:15487168

SOURCE HTG.

ORGANISM human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 176355)

Laird, G.

Direct Submission

Submitted (05-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerquest@sanger.ac.uk

On Sep 6, 2001 this sequence version replaced gi:14272282.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality &gt;=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TrEMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr10  
 RP11-453G10 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-453G10 It may be shorter because we sequence overlapping sections only once, except for a short overlap.  
 The true right end of clone RP11-453G10 is at 176355 in this sequence. The true right end of clone RP11-19501 is at 100 in this sequence.

## FEATURES

source

1. 176355  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-453G10"  
 /clone\_lib="RPCI-11.2"  
 6975. 17489  
 /note="match: GSS: Em:AQ730872"  
 6983. 17519  
 /note="match: GSS: Em:AQ727995"  
 30521. 30876  
 /note="match: GSS: Em:AQ210265"  
 complement(35673..36133)  
 /note="match: GSS: Em:AQ632919"  
 35866. 36208  
 /note="match: GSS: Em:AQ142257"  
 complement(77773..78285)  
 /note="match: GSS: Em:AQ149763"  
 complement(101941..102380)  
 /note="match: GSS: Em:AQ030914"  
 complement(118019..118511)  
 /note="match: GSS: Em:AQ628733"  
 complement(153629..153876)  
 /note="match: GSS: Em:AQ267194"  
 164173..164612  
 /note="match: GSS: Em:AQ133255"  
 171407..171956  
 /note="match: GSS: Em:AQ678987"

BASE COUNT

53311 a 33340 c 33627 g 56077 t

## ORIGIN

Query Match 91.1%; Score 16.4; DB 9; Length 176355;  
 Best Local Similarity 94.4%; Pred. No. 97;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atactctggctcatgttaa 18

|||||

Db 124306 ATACTCTGGTCATGTAA 124323

Search completed: July 31, 2002, 14:02:35  
 Job time: 17597 sec

us-09-899-718a-8.rge

Thu Aug 1 08:30:06 2002



|                          |        |               |       |              |
|--------------------------|--------|---------------|-------|--------------|
| Query Match              | 71.7%; | Score 17.2;   | DB 1; | Length 1996; |
| Best Local Similarity    | 86.4%; | Pred. No. 76; |       |              |
| Matches 19: Conservative | 0;     | Mismatches    | 3;    | Indels 0     |

RESULT 5  
US-08-303-861-1  
: Sequence 1, Application US/08303861

REMARKS 6  
US-09-060-410-3/c  
; Sequence 3, Application US/09060410  
; Patent No. 6165461  
; GENERAL INFORMATION:  
; APPLICANT: Cobb, Melanie  
; APPLICANT: Hutchinson, Michele  
; APPLICANT: Chen, Zhu  
; APPLICANT: Berman, Kevin  
; TITLE OF INVENTION: TAO PROTEIN KINASES AND METHODS OF USE  
; TITLE OF INVENTION: THEREFOR  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:

|                       |              |                         |               |             |
|-----------------------|--------------|-------------------------|---------------|-------------|
| Query Match           | 69.2%;       | Score 16.6;             | DB 4;         | Length 374; |
| Best Local Similarity | 79.2%;       | Pred. No. 1.3e+02;      |               |             |
| Matches 19;           | Conservative | 0;                      | Mismatches 5; | Indels 0;   |
| 0y                    | 1            | tcagccaggtccaccgcgtgcac | 24            |             |
|                       |              |                         |               |             |
| 0b                    | 101          | TCGCCAGNCTCCACCGGGCCACG | 78            |             |

us-09-899-718a-7.rni

Thu Aug 1 08:30:05 2002

```

;
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; TITLE OF INVENTION: ACETYL-CoA CARBOXYLASE AND USES
; THEREFOR
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/611,107
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US SN 07/956,700
; FILING DATE: 02-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/09340
; FILING DATE: 30-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US SN 08/422,560
; FILING DATE: 14-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: ARCD:221
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1458 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-611-107-1

Query Match 69.2%; Score 16.6; DB 1; Length 1458;
Best Local Similarity 82.6%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcagccaggtccaccctgac 23
Db 387 TTAGCCAGTTCACCTTAGCAC 365

RESULT 10
US-08-468-793-1/c
; Sequence 1, Application US/08468793
; Patent No. 6177267
; GENERAL INFORMATION:
; APPLICANT: Haselkorn, Robert
; APPLICANT: Gornicki, Piotr
; TITLE OF INVENTION: ACETYL-CoA CARBOXYLASE COMPOSITIONS AND
; TITLE OF INVENTION: METHODS OF USE
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,793
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,560
; FILING DATE: 14-APR-1995
; APPLICATION NUMBER: US SN 07/956,700
; FILING DATE: 02-OCT-1992
; ADDRESSEE: Arnold, White & Durkee

;
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/422,560A
; FILING DATE: 14-APR-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/956,700
; FILING DATE: 02-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mark B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: ARCD:152/WIM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512-418-3000
; TELEFAX: 512-474-7577
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1458 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-422-560A-1

Query Match 69.2%; Score 16.6; DB 2; Length 1458;
Best Local Similarity 82.6%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcagccaggtccaccctgac 23
Db 387 TTAGCCAGTTCACCTTAGCAC 365

RESULT 10
US-08-422-560A-1/c
; Sequence 1, Application US/08422560A
; Patent No. 5910626
; GENERAL INFORMATION:
; APPLICANT: Haselkorn, Robert
; APPLICANT: Gornicki, Piotr
; TITLE OF INVENTION: ACETYL-CoA CARBOXYLASE COMPOSITIONS AND
; TITLE OF INVENTION: METHODS FOR USE
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
```

CLASSIFICATION: 800  
APPLICATION NUMBER: PCT/US93/09340  
FILING DATE: 30-SEP-1993  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Kitchell, Barbara S.  
REGISTRATION NUMBER: 33,928  
REFERENCE/DOCKET NUMBER: ARCD:152/KIT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1458 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
US-08-468-793-1

Query Match 69.2%; Score 16.6; DB 4; Length 1458;  
Best Local Similarity 82.6%; Pred. No. 1.3e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcagccaggttcaccccgtagc 23  
Db 387 ttaccaggttcaccccgtagc 365

RESULT 11  
US-08-581-148C-15  
Sequence 15, Application US/08581148C  
Patent No. 6060644  
GENERAL INFORMATION:  
APPLICANT: Schnable, Patrick S.  
APPLICANT: Robertson, Donald S.  
APPLICANT: Hansen, Joel D.  
APPLICANT: Nikolau, Basil J.  
APPLICANT: Xu, Xiaojie  
TITLE OF INVENTION: ISOLATION AND USE OF CUTICULAR LIPID  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leydig, Voit & Mayer, Ltd.  
STREET: Two Prudential plaza, Suite 4900  
CITY: Chicago  
STATE: IL  
COUNTRY: US  
ZIP: 60601-6780  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/581,148C  
FILING DATE: 29-DEC-1995  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Larcher, Carol  
REGISTRATION NUMBER: 35243  
REFERENCE/DOCKET NUMBER: 71380  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 616-5600  
TELEFAX: (312) 616-5700  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1903 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double

TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1665  
US-08-581-148C-15

Query Match 69.2%; Score 16.6; DB 3; Length 1903;  
Best Local Similarity 82.6%; Pred. No. 1.3e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcagccaggttcaccccgtagc 23  
Db 266 TCAGCCGTTACCACTCCTTGAC 288

RESULT 12  
US-07-903-029-2/c  
Sequence 2, Application US/07903029  
Patent No. 5969097  
GENERAL INFORMATION:  
APPLICANT: Wiegand, Roger C.  
APPLICANT: Currie, Mark C.  
APPLICANT: Fok, Kam F.  
TITLE OF INVENTION: Human Guanylin  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dennis A. Bennett, Monsanto Co., A3SG  
STREET: 800 N. Lindbergh Blvd.  
CITY: St. Louis  
STATE: Missouri  
COUNTRY: USA  
ZIP: 63167  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/903,029  
FILING DATE: 19920623  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Bennett, Dennis A.  
REGISTRATION NUMBER: 34,547  
REFERENCE/DOCKET NUMBER: 07-21(872)A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (314)694-5402  
TELEFAX: (314)694-9009  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 589 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-07-903-029-2

Query Match 67.5%; Score 16.2; DB 2; Length 589;  
Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcagccaggttcaccccgtagc 21  
Db 500 TCACCCAGTTCTCCCGGGC 480

RESULT 13  
US-09-155-768-1/c  
Sequence 1, Application US/09155768A  
Patent No. 6162908

us-09-899-718a-7.rni

Thu Aug 1 08:30:05 2002

GENERAL INFORMATION:  
APPLICANT: SEIKAGAKU CORPORATION  
TITLE OF INVENTION: POLYPEPTIDE OF HYALURONAN SYNTHASE OF  
HUMAN-ORIGIN AND DNA ENCODING THE SAME  
FILE REFERENCE: TOYAM37\_001APC  
CURRENT APPLICATION NUMBER: US/09/155,768A  
CURRENT FILING DATE: 1998-10-05  
EARLIER APPLICATION NUMBER: JP 8-084326  
EARLIER FILING DATE: 1996-04-05  
EARLIER APPLICATION NUMBER: JP 8-109663  
EARLIER FILING DATE: 1996-04-30  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 2117  
TYPE: DNA  
ORGANISM: HUMAN  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (149)...(1777)  
US-09-155-768-1

Query Match 67.5%; Score 16.2; DB 4; Length 2117;  
Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 gccagttccaccccggtgcacg 24  
||||| ||||| || |||||  
DB 1374 GCCAGTGGCAGCCCTGCACG 1354

RESULT 14  
US-09-103-840A-2/c  
Sequence 2, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 4403765  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
OTHER INFORMATION: CDC 1551  
OTHER INFORMATION: "n" bases at various positions throughout the sequence  
OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 67.5%; Score 16.2; DB 4; Length 4403765;  
Best Local Similarity 90.0%; Pred. No. 54; Mismatches 2; Indels 0; Gaps 0;  
Matches 18; Conservative 0

QY 2 cagccagttccaccccggtgc 21  
||||| ||||| ||||| |||||  
DB 2098669 CACCCAGTTCACCCCGGTGC 2098650

RESULT 15  
US-09-103-840A-1/c  
Sequence 1, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37Rv  
US-09-103-840A-1

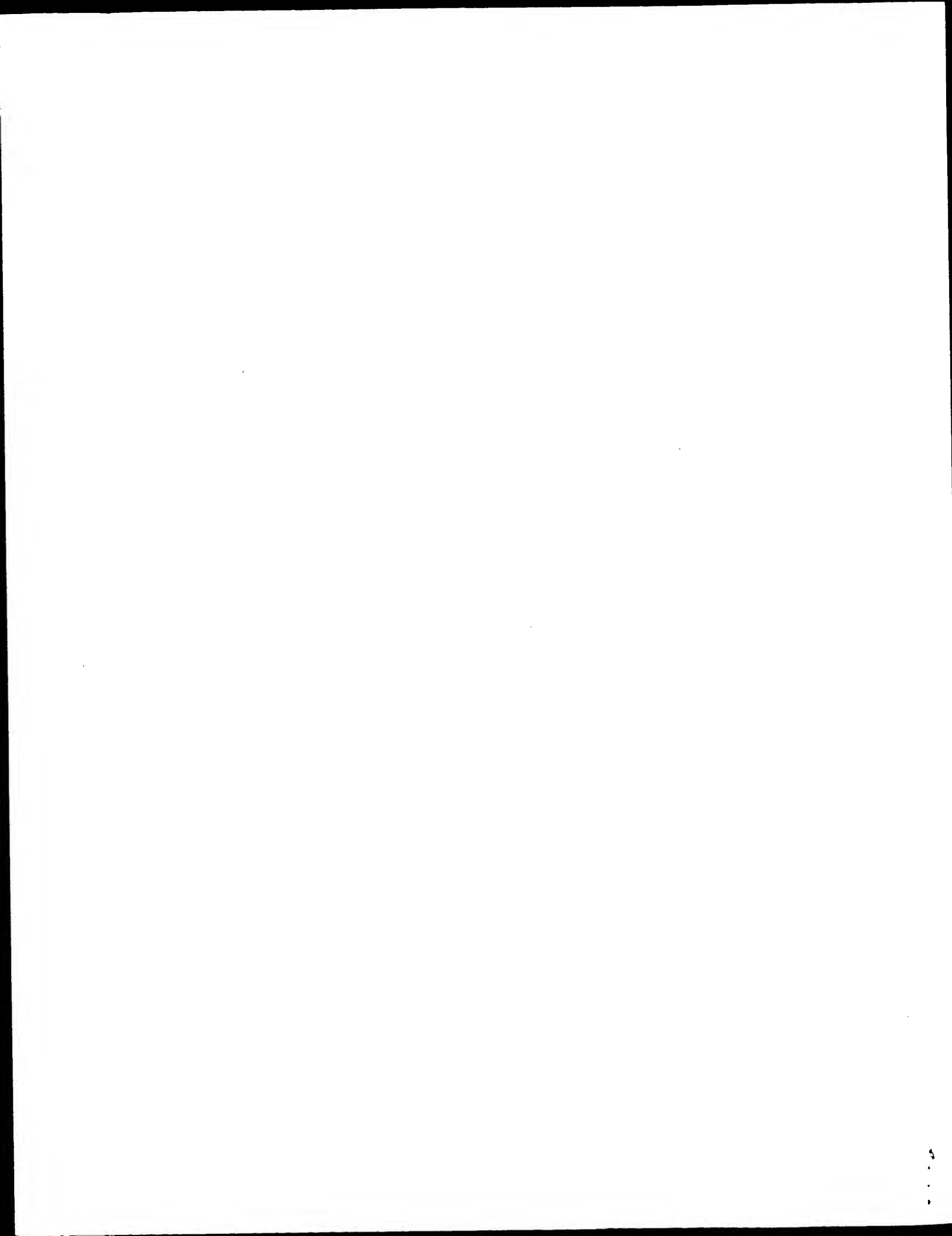
Query Match 67.5%; Score 16.2; DB 4; Length 4411529;  
Best Local Similarity 90.0%; Pred. No. 54; Mismatches 2; Indels 0; Gaps 0;  
Matches 18; Conservative 0

QY 2 cagccagttccaccccggtgc 21  
||||| ||||| ||||| |||||  
DB 2101479 CACCCAGTTCACCCCGGTGC 2101460

Search completed: July 31, 2002, 12:26:27  
Job time: 12010 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:12:09 ; Search time 5855.71 Seconds  
(without alignments)  
55.318 Million cell updates/sec

Title: US-09-899-718A-7  
Perfect score: 24  
Sequence: 1 tcagccagttccaccgcgtgcacg 24

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_gss.\*  
13: em\_gss\_hum.\*  
14: em\_gss\_inv.\*  
15: em\_gss\_pln.\*  
16: em\_gss\_vrt.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| C 1        | 19.4  | 80.8        | 836    | 10    | BG444193 GA_Ea002  |
| C 2        | 19.2  | 80.0        | 808    | 12    | AL288186 Tetraodon |
| C 3        | 18.8  | 78.3        | 420    | 10    | C19505 C19505 Rice |
| 4          | 18.8  | 78.3        | 453    | 9     | AU065405 AU065405  |
| 5          | 18.8  | 78.3        | 460    | 9     | AU065404 AU065404  |
| 6          | 18.2  | 75.8        | 306    | 10    | B1510068 B1510068  |
| C 7        | 18.2  | 75.8        | 426    | 10    | BG981373 CM4-CN008 |
| 8          | 18.2  | 75.8        | 528    | 9     | AL162491 A018F380  |
| 9          | 18.2  | 75.8        | 559    | 10    | BF191609 239242 MA |
| C 10       | 18.2  | 75.8        | 687    | 10    | BG827582 602748458 |
| C 11       | 18.2  | 75.8        | 702    | 10    | B1329586 602984002 |
| C 12       | 18.2  | 75.8        | 765    | 12    | BH045197 RPI-24-3  |
| C 13       | 18.2  | 75.8        | 907    | 10    | BG104185 602310751 |
| C 14       | 18.2  | 75.8        | 1200   | 11    | AK009013 Mus muscu |
| C 15       | 17.8  | 74.2        | 64     | 12    | AZ875573 2M0190602 |
| C 16       | 17.8  | 74.2        | 449    | 10    | BF739541 NCMI0C12T |
| 17         | 17.8  | 74.2        | 642    | 12    | AZ831095 2M0110112 |

|      |      |      |      |    |          |
|------|------|------|------|----|----------|
| C 18 | 17.8 | 74.2 | 768  | 10 | BG962250 |
| 19   | 17.8 | 74.2 | 1021 | 12 | CNS03RVF |
| 20   | 17.6 | 73.3 | 276  | 10 | B1527986 |
| 21   | 17.6 | 73.3 | 286  | 9  | A1104620 |
| 22   | 17.6 | 73.3 | 325  | 10 | BF415383 |
| 23   | 17.6 | 73.3 | 327  | 9  | AI639268 |
| 24   | 17.6 | 73.3 | 340  | 9  | A1414355 |
| C 25 | 17.6 | 73.3 | 346  | 9  | AV012778 |
| C 26 | 17.6 | 73.3 | 373  | 12 | AZ790364 |
| 27   | 17.6 | 73.3 | 385  | 12 | AZ790223 |
| 28   | 17.6 | 73.3 | 386  | 10 | BF407002 |
| 29   | 17.6 | 73.3 | 393  | 9  | AW523973 |
| 30   | 17.6 | 73.3 | 400  | 10 | B195885  |
| 31   | 17.6 | 73.3 | 413  | 9  | AW433934 |
| 32   | 17.6 | 73.3 | 416  | 9  | AI502941 |
| 33   | 17.6 | 73.3 | 460  | 9  | AW434970 |
| C 34 | 17.6 | 73.3 | 487  | 9  | BB857929 |
| 35   | 17.6 | 73.3 | 519  | 10 | BM426700 |
| 36   | 17.6 | 73.3 | 522  | 10 | B1993715 |
| 37   | 17.6 | 73.3 | 523  | 9  | AA816166 |
| C 38 | 17.6 | 73.3 | 524  | 9  | AI391290 |
| 39   | 17.6 | 73.3 | 542  | 9  | AI454923 |
| 40   | 17.6 | 73.3 | 569  | 10 | BG373840 |
| 41   | 17.6 | 73.3 | 579  | 9  | AV932786 |
| 42   | 17.6 | 73.3 | 716  | 10 | BG843380 |
| C 43 | 17.6 | 73.3 | 733  | 10 | BE566538 |
| C 44 | 17.6 | 73.3 | 769  | 10 | BF627856 |
| C 45 | 17.6 | 73.3 | 806  | 10 | BG334627 |

## ALIGNMENTS

RESULT 1  
BG444193/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BG444193 836 bp mRNA linear EST 15-MAR-2001  
GA\_Ea0023J06f Gossypium arboreum 7-10 dpa fiber library Gossypium  
arboreum cDNA clone GA\_Ea0023J06f, mRNA sequence.  
BG444193  
BG444193.1 GI:13353845  
EST.  
Gossypium arboreum.  
Gossypium arboreum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.  
1 (bases 1 to 836)  
Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry  
D., Wood,T.C., Leslie,A. and Wilkins,T.A.  
An integrated analysis of the genetics, development, and evolution  
of the cotton fiber  
Unpublished (2000)  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: TAATACGACTCATATAGGG  
High quality sequence start: 2  
High quality sequence stop: 826.  
Location/Qualifiers  
1. 836  
/organism="Gossypium arboreum"  
/strain="ARA"  
/cultivar="8400"  
/db\_xref="taxon:29729"  
/clone="GA\_Ea0023J06f"  
/tissue\_type="Fibers isolated from bolls harvested 7-10  
dpa"  
/lab\_host="E. coli"

FEATURES  
source

Thu Aug 1 08:30:06 2002

```

C19505
LOCUS          C19505 Rice panicle at ripening stage Oryza sativa cDNA clone
DEFINITION     E10525_1A, mRNA sequence.
ACCESSION      C19505.1 GI:1631776
VERSION        Oryza sativa.
KEYWORDS       EST.
SOURCE         Oryza sativa.
ORGANISM       Oryza sativa.
REFERENCE      Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
                Ehrhartoideae; Oryzeae; Oryza.
                1 (bases 1 to 420)
AUTHORS        Sasaki,T.
TITLE          Rice cDNA from panicle at ripening stage
JOURNAL        Unpublished (1996)
COMMENT        Contact: Takuji Sasaki
                National Institute of Agrobiological Resources
                Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
                305-8602, Japan
                Tel: 81-298-38-7441
                Fax: 81-298-38-7468
                Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/.

FEATURES             source
     1..420
     /organism="Oryza sativa"
     /strain="Nipponbare"
     /db_xref="taxon:4530"
     /clone="E10525_1A"
     /clone.lib="Rice panicle at ripening stage"
     /dev_stage="ripening stage"
     /note="Organ: panicle; Rice cDNA from panicle at ripening
     stage"
BASE COUNT      92 a 132 c 121 g 71 t 4 others
ORIGIN

Query Match      78.3%; Score 18.8; DB 10; Length 420;
Best Local Similarity 90.9%; Pred. No. 1.3e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tcagccagttccaccctgcac 22
    |||||
Db 323 TCGCCAGTTCACCCCGTGAA 344

RESULT 4
LOCUS          AU065405 Rice root Oryza sativa cDNA clone R3592_1A, mRNA sequence.
DEFINITION     AU065405
ACCESSION      AU065405
VERSION        AU065405.1 GI:4881106
KEYWORDS       EST.
SOURCE         Oryza sativa.
ORGANISM       Oryza sativa.
REFERENCE      Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
                Ehrhartoideae; Oryzeae; Oryza.
                1 (bases 1 to 453)
AUTHORS        Minobe,Y. and Sasaki,T.
TITLE          Rice cDNA from root
JOURNAL        Unpublished (1995)
COMMENT        Contact: Takuji Sasaki
                National Institute of Agrobiological Resources
                Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
                305-8602, Japan
                Tel: 81-298-38-7441
                Fax: 81-298-38-7468
                Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
                PROJECT "RGP"
                Location/Qualifiers
                1..453
                /organism="Oryza sativa"

FEATURES             source
     1..453
     /organism="Oryza sativa"

```

```

C19505
LOCUS          C19505 Rice panicle at ripening stage Oryza sativa cDNA clone
DEFINITION     E10525_1A, mRNA sequence.
ACCESSION      C19505.1 GI:1631776
VERSION        Oryza sativa.
KEYWORDS       EST.
SOURCE         Oryza sativa.
ORGANISM       Oryza sativa.
REFERENCE      Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
                Ehrhartoideae; Oryzeae; Oryza.
                1 (bases 1 to 420)
AUTHORS        Sasaki,T.
TITLE          Rice cDNA from panicle at ripening stage
JOURNAL        Unpublished (1996)
COMMENT        Contact: Takuji Sasaki
                National Institute of Agrobiological Resources
                Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
                305-8602, Japan
                Tel: 81-298-38-7441
                Fax: 81-298-38-7468
                Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/.

FEATURES             source
     1..420
     /organism="Oryza sativa"
     /strain="Nipponbare"
     /db_xref="taxon:4530"
     /clone="E10525_1A"
     /clone.lib="Rice panicle at ripening stage"
     /dev_stage="ripening stage"
     /note="Organ: panicle; Rice cDNA from panicle at ripening
     stage"
BASE COUNT      92 a 132 c 121 g 71 t 4 others
ORIGIN

Query Match      78.3%; Score 18.8; DB 10; Length 420;
Best Local Similarity 90.9%; Pred. No. 1.3e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tcagccagttccaccctgcac 22
    |||||
Db 323 TCGCCAGTTCACCCCGTGAA 344

RESULT 4
LOCUS          AU065405 Rice root Oryza sativa cDNA clone R3592_1A, mRNA sequence.
DEFINITION     AU065405
ACCESSION      AU065405
VERSION        AU065405.1 GI:4881106
KEYWORDS       EST.
SOURCE         Oryza sativa.
ORGANISM       Oryza sativa.
REFERENCE      Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
                Ehrhartoideae; Oryzeae; Oryza.
                1 (bases 1 to 453)
AUTHORS        Minobe,Y. and Sasaki,T.
TITLE          Rice cDNA from root
JOURNAL        Unpublished (1995)
COMMENT        Contact: Takuji Sasaki
                National Institute of Agrobiological Resources
                Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
                305-8602, Japan
                Tel: 81-298-38-7441
                Fax: 81-298-38-7468
                Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
                PROJECT "RGP"
                Location/Qualifiers
                1..453
                /organism="Oryza sativa"

FEATURES             source
     1..453
     /organism="Oryza sativa"

```

/strain="Nipponbare, sub\_species Japonica"  
 /db\_xref="taxon:4530"  
 /clone="R3592\_1A"  
 /clone\_lib="Rice root"  
 /note="Prepared from seedling root." 4 others  
 BASE COUNT 103 a 136 c 126 g 84 t  
 ORIGIN

Query Match 78.3%; Score 18.8; DB 9; Length 453;  
 Best Local Similarity 90.9%; Pred. No. 1.3e+03;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 tcagccagttccaccctgtgca 22  
 || |||||  
 Db 325 TCGGCCAGTTCACCCCGTGAA 346

RESULT 5  
 AU065404  
 LOCUS  
 DEFINITION AU065404 Rice root Oryza sativa cdna clone R3590\_1A, mRNA sequence. EST 20-MAY-1999  
 ACCESSION AU065404  
 VERSION AU065404.1 GI:4881105  
 KEYWORDS EST.  
 SOURCE  
 ORGANISM Oryza sativa.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 1 (bases 1 to 460)  
 Minobe, Y. and Sasaki, T.  
 Rice cdna from root  
 Unpublished (1995)  
 CONTACT: Takuji Sasaki  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki  
 305-8602, Japan  
 Tel: 81-298-38-7441  
 Fax: 81-298-38-7468  
 Email: tsasaki@abrr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/  
 PROJECT = "RGP".

FEATURES  
 source  
 1..460  
 /organism="Oryza sativa"  
 /strain="Nipponbare, sub\_species Japonica"  
 /db\_xref="taxon:4530"  
 /clone="R3590\_1A"  
 /clone\_lib="Rice root"  
 /note="Prepared from seedling root." 4 others  
 BASE COUNT 103 a 145 c 128 g 80 t  
 ORIGIN

Query Match 78.3%; Score 18.8; DB 9; Length 460;  
 Best Local Similarity 90.9%; Pred. No. 1.3e+03;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 tcagccagttccaccctgtgca 22  
 || |||||  
 Db 328 TCGGCCAGTTCACCCCGTGAA 349

RESULT 6  
 BI510068  
 LOCUS  
 DEFINITION BI510068 BBI70026B10G05 Bee Brain Normalized/Subtracted Library, BBI7 Apis EST 29-AUG-2001  
 ACCESSION BI510068  
 VERSION BI510068.1 GI:15360442  
 KEYWORDS EST.  
 SOURCE honeybee.  
 ORGANISM Apis mellifera

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata  
 ; Apoidea; Apidae; Apis.  
 1 (bases 1 to 306)  
 Whitfield, C.W., Soares, B., Robertson, H.M., Pardinas, J., Liu, L.,  
 Smoller, D. and Robinson, G.E.  
 An Expressed Sequence Tag Resource for Studies of Brain and  
 Behavior in the Honey Bee  
 Unpublished (2001)  
 CONTACT: Gene E. Robinson  
 Department of Entomology  
 University of Illinois  
 505 S. Goodwin Ave., Urbana, IL 61801, USA  
 Tel: 217 265 0309  
 Fax: 217 244 3499  
 Email: generobi@life.uiuc.edu  
 This research was funded by the University of Illinois Critical  
 Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation  
 Award in Functional Genomics to G.E. Robinson and an NSF  
 Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.  
 PCR PRIMERS  
 FORWARD: TAATACGACTCCTACTATAGG  
 BACKWARD: ATTAACCCCTCACTAAAG  
 Insert Length: 306 Std Error: 0.00  
 Plate: BBI70026B10 row: G column: 05  
 Seq primer: AGCGGATAACAATTTCACACAGGA  
 High quality sequence stop: 306.  
 Location/Qualifiers  
 1..306  
 /organism="Apis mellifera"  
 /strain="mixed strains of European bees, predominantly  
 A.m. ligustica"  
 /db\_xref="taxon:7460"  
 /clone="BBI70026B10G05"  
 /clone\_lib="Bee Brain Normalized/Subtracted Library, BBI7"  
 /sex="female"  
 /tissue\_type="brain"  
 /dev\_stage="adult worker honey bee"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pT73-Pac; Site\_1: ECORI;  
 Site\_2: NotI; This BBI7 cdna library was generated by  
 subtraction of the BBI6 library with 4000 previously  
 sequenced clones. The BBI6 library was contributed by the  
 Soares laboratory and it was constructed and normalized by the  
 as described by Ronaldo, M.F., Lennon, G. and Soares,  
 M.B. (1996), Genome Research 6(9): 791-806. RNA was  
 prepared from dissected brains of adult worker bees of  
 various ages and various behavioral groups."

## BASE COUNT

94 a 67 c 59 g 86 t

## ORIGIN

Query Match 75.8%; Score 18.2; DB 10; Length 306;  
 Best Local Similarity 87.0%; Pred. No. 2.2e+03;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 tcagccagttccaccctgtgca 23  
 || |||||  
 Db 256 TCGGCCAGTTCACCCCGTGCTC 278

## RESULT 7

BG981373/c

## LOCUS

BG981373

CM4-CN0089-070201-746-e01 CN0089 Homo sapiens cdna, mRNA sequence. EST.

ACCESSION BG981373

VERSION BG981373.1 GI:14384108

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 426)

Thu Aug 1 08:30:06 2002

**AUTHORS**  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

**TITLE**  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

**JOURNAL MEDLINE COMMENT**  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM4st2-CM4-CN0089-070201-746-e01st3-2001-02-07st4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 54  
High quality sequence stop: 426.  
Location/Qualifiers  
1. 426  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CN0089"  
/dev\_stage="Adult"  
/note="Organ: colon\_normal; Vector: puc18; Site.1: Smal; Site.2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (O.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
BASE COUNT 85 a 131 c 103 g 107 t  
ORIGIN

Query Match 75.8%; Score 18.2; DB 10; Length 426;  
Best Local Similarity 87.08; Pred. No. 2.3e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcagccagttccaccgccgtgcac 23  
|||||  
Db 299 TCACAGAGTTCACCCCATACAC 277  
|||||

**RESULT**  
8  
AL162491  
LOCUS A018P38U Hybrid aspen plasmid library Populus tremula x Populus tremuloides cDNA 5', mRNA sequence.  
DEFINITION  
ACCESSION AL162491.1 GI:3853776  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Populus tremula x Populus tremuloides.  
Populus tremula x Populus tremuloides.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucosids I; Malpighiales; Salicaceae; Populus.  
1 (bases 1 to 528)  
Sterky,F., Regan,S., Karlsson,J., Hertzberg,M., Rohde,A., Holmberg,A., Amiri,B., Bhalarao,R., Larsson,M., Villarroel,R., Van Montagu,M., Sandberg,G., Olsson,O., Teeri,T.T., Boerjan,W., Gustafsson,P., Uhlen,M., Sundberg,B. and Lundberg,J.  
Gene discovery in the wood-forming tissues of poplar: Analysis of 5,692 expressed sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13330-13335 (1998)  
99007314

**AUTHORS**  
Contact: Sterky F  
Department of Biotechnology  
Royal Institute of Technology (KTH)  
Teknikringen 34, S-100 44 STOCKHOLM, Sweden  
Tel: +46 8 790 8287  
Fax: +46 8 24 54 52  
Email: fredrik@biochem.kth.se  
PCR Primers  
FORWARD: AAGAGGGGATGCTGTCGAAGCGG  
BACKWARD: GCTTCGGCTGCTATGTTGTG  
Seq primer: CTTGTAAACGACGCCAG  
High quality sequence stop: 528.  
Location/Qualifiers  
1. 528  
/organism="Populus tremula x Populus tremuloides"  
/db\_xref="taxon:47664"  
/clone\_lib="Hybrid aspen plasmid library"  
/tissue\_type="Cambial region"  
/dev\_stage="1.5 m actively growing tree"  
/lab\_host="E.coli"  
/note="vector: pBluescript SK; Site.1: SalI; Site.2: NotI; Cambial region tissues, including developing xylem, the meristematic cambial zone and the developing and mature phloem, was harvested from 1.5 m actively growing trees. cDNA was prepared and cloned into lambda gt22a. DNA was isolated and subcloned into pBluescript SK using SalI and NotI restriction enzymes."  
BASE COUNT 116 a 174 c 111 g 124 t 3 others  
ORIGIN

Query Match 75.8%; Score 18.2; DB 9; Length 528;  
Best Local Similarity 87.08; Pred. No. 2.3e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcagccagttccaccgccgtgcac 23  
|||||  
Db 398 TCAGCAGTTCACCCCGGTGCAC 420  
|||||

**RESULT**  
9  
BF191609  
LOCUS 239242 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.  
DEFINITION  
ACCESSION BF191609  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
1 (bases 1 to 559)  
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W. and Keele,J.W.  
Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine  
EST discovered (2000)  
Unpublished (2000)  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@mail.marc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -minmatch 12 options.  
PCR Primers  
FORWARD: AGGAACAGCATGACCAT  
BACKWARD: GTTTCCCGACGACGACG  
Plate: 67 row: B column: 7  
Seq primer: ATTTAGCTGACACTAG.  
Location/Qualifiers

```

source
1..559
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2PIC"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT 117 a 181 c 152 g 109 t
ORIGIN

Query Match 75.8%; Score 18.2; DB 10; Length 559;
Best Local Similarity 87.0%; Pred. No. 2.4e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tcagccaggttcaccccggtgcac 23
||||| ||||| |||||
Db 269 TCAGCCAGATCCACCCATGCAC 291

RESULT 10
BG827582/c
LOCUS
DEFINITION 602748468F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4901416 5',
mRNA sequence.
ACCESSION BG827582
VERSION BG827582.1 GI:14175169
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 687)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1796 row: g column: 17
High quality sequence start: 4
High quality sequence stop: 551.
Location/Qualifiers
1..687
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4901416"
/clone_lib="NIH_MGC_17"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: muscle; Vector: pONB7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dt priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 162 a 178 c 224 g 123 t
ORIGIN

Query Match 75.8%; Score 18.2; DB 10; Length 687;
Best Local Similarity 87.0%; Pred. No. 2.4e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

source
1..587
/organism="Mus musculus"
/db_xref="FVB/N"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5136697"
/clone_lib="NCI_CGAP_L19"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 130 a 237 c 204 g 131 t
ORIGIN

Query Match 75.8%; Score 18.2; DB 10; Length 702;
Best Local Similarity 87.0%; Pred. No. 2.4e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tcagccaggttcaccccggtgcac 23
||||| ||||| |||||
Db 644 TCAGCCAGTTCACCCCTGTGCAC 622

RESULT 12
BH045197/c
LOCUS
DEFINITION BH045197 RPCI-24-301P3-TV RPCI-24 Mus musculus genomic clone RPCI-24-301P3,
DNA sequence.
ACCESSION BH045197
VERSION BH045197.1 GI:14829258
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 765)
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,
Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E.,

```

100



KEYWORDS  
SOURCE  
ORGANISM  
  
REFERENCE  
AUTHORS  
  
TITLE  
JOURNAL  
COMMENT  
  
FEATURES  
source

[illegible]

```

laboratory mouse DNA resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (g1147321141gb1AF129072.1), a copy-number
inducible derivative of plasmid RI. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

16 a      20 c      13 g      15 t
ORIGIN

Query Match          74.2%;   Score 17.8;   DB 12;   Length 64;
Best Local Similarity 90.5%;   Pred. No. 2.5e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2   cagccagttccaccctgca 22
          |||||
Db      34   CAGCCAGTTCACACCCGTGCA 54

Search completed: July 31, 2002, 12:12:12
Job time: 13485 sec

```

Search completed: July 31, 2002, 12:12:12  
Job time: 13485 sec

us-09-899-718a-7.rst

Thu Aug 1 08:30:06 2002

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:11:24 ; Search time 720.8 seconds  
(without alignments)  
57.167 Million cell updates/sec

Title: US-09-899-718a-7  
Perfect score: 24  
Sequence: 1 tcagcagttccaccggtgcag 24

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues  
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID       | Description        |
|------------|-------|-------------|--------|-------------|--------------------|
| 1          | 18.2  | 75.8        | 7555   | 11 AAQ05831 | Cardiac sodium cha |
| 2          | 18.2  | 75.8        | 7555   | 16 AAQ81328 | Cardiac sodium cha |
| 3          | 17.6  | 73.3        | 57     | 15 AAQ73770 | Hepatitis C virus  |
| 4          | 17.4  | 72.5        | 4541   | 23 AB124072 | Drosophila melanog |
| 5          | 17.2  | 71.7        | 230    | 22 AAH27139 | Human beta-2 adren |
| 6          | 17.2  | 71.7        | 657    | 15 AAQ54681 | Rat proteosome RIN |
| 7          | 17.2  | 71.7        | 1785   | 24 AAS19511 | Reference sequence |
| 8          | 17.2  | 71.7        | 1972   | 23 AAS91810 | DNA encoding novel |
| 9          | 17.2  | 71.7        | 1996   | 18 AAH90830 | cDNA encoding gluc |

|      |      |      |       |    |          |                     |
|------|------|------|-------|----|----------|---------------------|
| c 10 | 17.2 | 71.7 | 2300  | 20 | AAH61116 | Human beta2-adrene  |
| c 11 | 17.2 | 71.7 | 2305  | 21 | AAA38340 | Human beta-adrener  |
| c 12 | 17.2 | 71.7 | 2340  | 21 | AAA38784 | Human beta2 adrene  |
| c 13 | 17.2 | 71.7 | 2745  | 18 | AAH90831 | Glucosylase P gen   |
| c 14 | 17.2 | 71.7 | 3451  | 19 | AAV52614 | Human beta-2-adren  |
| c 15 | 17.2 | 71.7 | 3451  | 20 | AAZ00774 | Human beta 2-adren  |
| c 16 | 17.2 | 71.7 | 3451  | 20 | AAZ00775 | Human beta 2-adren  |
| c 17 | 17.2 | 71.7 | 3451  | 20 | AAZ00776 | Human beta 2-adren  |
| c 18 | 17.2 | 71.7 | 3451  | 20 | AAZ00777 | Human beta 2-adren  |
| c 19 | 17.2 | 71.7 | 3451  | 20 | AAZ00778 | Human beta 2-adren  |
| c 20 | 17.2 | 71.7 | 3451  | 20 | AAZ00779 | Human beta 2-adren  |
| c 21 | 17.2 | 71.7 | 3451  | 20 | AAZ00780 | Human beta 2-adren  |
| c 22 | 17.2 | 71.7 | 3451  | 20 | AAZ00773 | Human beta 2-adren  |
| c 23 | 17.2 | 71.7 | 3451  | 21 | AAA38339 | Human beta-adrener  |
| c 24 | 17.2 | 71.7 | 3451  | 24 | AAH18444 | Reference sequence  |
| c 25 | 17.2 | 71.7 | 4576  | 23 | AAH89475 | DNA encoding novel  |
| c 26 | 17.2 | 71.7 | 11580 | 24 | ABA92604 | Human reelin encod  |
| c 27 | 17.2 | 71.7 | 11632 | 23 | AAH89484 | DNA encoding novel  |
| c 28 | 17.2 | 71.7 | 32768 | 20 | AAH20515 | Polynucleotide seq  |
| c 29 | 17.2 | 70.8 | 411   | 22 | AAH61727 | Human immune/haema  |
| c 30 | 16.8 | 70.0 | 230   | 16 | AAQ83564 | Renal carcinoma as  |
| c 31 | 16.8 | 70.0 | 320   | 22 | ABA47238 | Human breast cell   |
| c 32 | 16.8 | 70.0 | 320   | 22 | ABA65123 | Human foetal liver  |
| c 33 | 16.8 | 70.0 | 320   | 22 | ABA32226 | Probe #10692 for g  |
| c 34 | 16.8 | 70.0 | 320   | 22 | AAH13542 | Human brain expres  |
| c 35 | 16.8 | 70.0 | 320   | 22 | AAH39281 | Human bone marrow   |
| c 36 | 16.8 | 70.0 | 320   | 22 | AAH20094 | Probe #10027 for g  |
| c 37 | 16.8 | 70.0 | 320   | 22 | AAH45293 | Probe #13979 used t |
| c 38 | 16.8 | 70.0 | 320   | 22 | AAH05800 | Probe #5791 used t  |
| c 39 | 16.8 | 70.0 | 466   | 20 | AAH87537 | EST clone DD401.    |
| c 40 | 16.8 | 70.0 | 493   | 22 | ABA42084 | Human breast cell   |
| c 41 | 16.8 | 70.0 | 493   | 22 | ABA52506 | Human foetal liver  |
| c 42 | 16.8 | 70.0 | 493   | 22 | ABA22296 | Probe #762 for gen  |
| c 43 | 16.8 | 70.0 | 493   | 22 | AAH00770 | Human brain expres  |
| c 44 | 16.8 | 70.0 | 493   | 22 | AAH26222 | Human bone marrow   |
| c 45 | 16.8 | 70.0 | 493   | 22 | AAH10855 | Probe #788 for gen  |

## ALIGNMENTS

RESULT 1  
AAQ05831  
ID AAQ05831 standard; cDNA; 7555 BP.  
AC AAQ05831;  
XX  
XX 10-JAN-1991 (first entry)  
DT  
DE Cardiac sodium channel gene.  
DE  
DE  
DE  
KW Rat; arrhythmia; ss.  
XX  
XX Rattus rattus.  
XX  
XX WO9009391-A.  
XX  
XX 23-AUG-1990.  
PD  
PF 09-FEB-1990; 90WO-US00768.  
XX  
XX 13-FEB-1989; 89US-0310330.  
PR  
PA (ARCH-) ARCH DEV CORP.  
XX  
XX Rogart RB;  
PI  
XX  
XX WPI: 1990-275095/36.  
DR  
DR P-PSDB; AAR06584.  
XX  
XX New rat cardiac sodium channel proteins - and associated DNA  
XX sequences, polypeptides and peptides associated with  
XX proteins, useful as antiarrhythmic and cardiotoxic drugs.

plasmid PRH3-1 (ATCC 67885), plasmid PRH4-23 (ATCC 67896) and plasmid PRH14-31 (ATCC 67887). A virus/circular DNA plasmid vector comprising the cDNA may be transformed or transfected into a prokaryote/eukaryote host cell, and the resulting recombinant sodium channel protein has various therapeutic, diagnostic and prognostic uses. It may also be used to develop more effective antiarrhythmic, cardiant and cardioglycoside drugs.

ABL24072/c  
ID ABL24072 standard; DNA; 4541 BP.

ABL24072/c  
ID ABL24072 standard; DNA; 4541 BP.

AC ABL24072;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 23689.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 XX 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 XX WPI; 2001-656860/75.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 23689; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (AB57737-AB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 4541 BP; 1232 A; 1097 C; 1013 G; 1199 T; 0 other;  
 SQ

Query Match 72.5%; Score 17.4; DB 23; Length 4541;  
 Best Local Similarity 94.7%; Pred. No. 2.7e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 agccagttccaccctgtgc 21  
 ||||| ||||| ||||| |||||  
 Db 1256 AGCCAAATTCACCCCGTGC 1238

RESULT 5  
 AAH27139/c  
 ID AAH27139 standard; DNA; 230 BP.  
 XX  
 AC AAH27139;  
 XX  
 DT 08-AUG-2001 (first entry)  
 XX  
 DE Human beta-2 adrenergic receptor UTR region with RBP binding ability.  
 XX  
 KW Untranslated region; UTR; RNA binding protein; RBP; neurodegeneration;  
 KW stroke; cardiovascular disease; hypertension; cancer; inflammation;  
 KW metabolic disorder; obesity; diabetes; beta-2 adrenergic receptor; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200134624-A1.

XX 17-MAY-2001.  
 XX  
 PF 09-NOV-2000; 2000WO-US30888.  
 XX  
 PR 10-NOV-1999; 99US-0437458.  
 XX  
 PA (MESS-) MESSAGE PHARM INC.  
 XX  
 PI Giordano A, Xavier AK;  
 XX  
 XX WPI; 2001-335904/35.  
 XX  
 XX New nucleic acids that bind RNA-binding proteins or regulate mRNA  
 PT function, useful for therapeutic gene regulation, such as in cases of  
 PT neurodegeneration -  
 XX  
 XX Claim 1; Page 28; 33pp; English.  
 XX  
 CC Sequences AAH27132 - AAH27151 represent human gene untranslated regions  
 CC where the corresponding mRNA fragment has RNA binding protein (RBP)  
 CC binding activity. RBPs mediate the processing of pre-mRNA, the transport  
 CC of mRNA from the nucleus to the cytoplasm, mRNA stabilisation, therefore  
 CC translational efficiency, and the sequestration of some mRNAs. Therefore  
 CC modification of post-transcriptional protein expression in eukaryotic  
 CC cells may be carried out through the targeting of specific interactions of  
 CC proteins that bind to RBPs. The gene fragments of the invention are used  
 CC to identify their optimized sub-fragments, compounds that affect RNA/RBP  
 CC interaction or mRNA functionality; or RBPs that interact with the  
 CC compounds. Compounds identified using the gene fragments are potentially  
 CC useful for therapeutic regulation of gene expression, such as in cases of  
 CC neurodegeneration; stroke; cardiovascular disease; hypertension; cancer;  
 CC inflammation; metabolic disorders (obesity and diabetes) and bacterial or  
 CC viral infection. The present sequence is one of gene fragments of the  
 CC invention, isolated from the human beta-2 adrenergic receptor gene.  
 XX  
 XX Sequence 230 BP; 42 A; 91 C; 70 G; 27 T; 0 other;  
 SQ

Query Match 71.7%; Score 17.2; DB 22; Length 230;  
 Best Local Similarity 86.4%; Pred. No. 2.7e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cagccagttccaccctgtgc 23  
 ||||| ||||| ||||| |||||  
 Db 42 CTGCCAGTTCCAGCCCGTGC 21

RESULT 6  
 AAQ54681/c  
 ID AAQ54681 standard; cDNA to mRNA; 657 BP.  
 XX  
 AC AAQ54681;  
 XX  
 DT 06-JUL-1994 (first entry)  
 XX  
 DE Rat proteosome RING12 gene.  
 XX  
 KW Component; rat; proteosome; vector; transformant; tumours;  
 KW diagnosis; expression vector; ss.  
 XX  
 OS Rattus rattus.  
 XX  
 PH Key Location/Qualifiers  
 FT CDS 1..657  
 FT /\*tag= a  
 FT /product= rat\_proteosome\_RING12  
 XX  
 PN JP05317059-A.  
 XX  
 XX 03-DEC-1993.  
 PD  
 XX 22-MAY-1992; 92JP-0154184.  
 PF

XX 22-MAY-1992; 92JP-0154184.  
 XX PA (BIOM-) BIO MATERIAL KENYUSHO KK.  
 XX WPI: 1994-011024/02.  
 DR P-PSDB; AAR47473.  
 XX  
 XX A component of rat proteasome - for elucidation of mechanism of  
 PT e.g. malignant tumour  
 PT  
 XX Claim 8; Page 5-6; 24pp; Japanese.  
 PS  
 CC Sequences (AA054678-81) show various components of a rat proteasome.  
 CC The component is useful for the elucidation of various mechanisms  
 CC of diseases such as malignant tumours and their diagnosis and  
 CC treatment.  
 CC  
 XX Sequence 657 BP; 137 A; 184 C; 198 G; 138 T; 0 other;  
 SQ

Query Match 71.7%; Score 17.2; DB 15; Length 657;  
 Best Local Similarity 86.4%; Pred. No. 2.9e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcagccagttccaccctgtca 22  
 || ||||| ||||| |||||  
 Db 275 TCCTCCAGTTCACCGGTGCA 254

RESULT 7  
 AAS19511  
 ID AAS19511 standard; DNA; 1785 BP.  
 XX AC AAS19511;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Reference sequence for human MPL gene exons 9-10.  
 XX  
 KW Human; single nucleotide polymorphism; SNP; MPL; chromosome 1p34;  
 KW myeloproliferative leukaemia virus oncogene; haplotyping; genotyping;  
 KW congenital megakaryocytic thrombocytopaenia; CAMT; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT exon 614..773  
 FT /\*tag= a  
 FT /\*number= 9  
 FT intron 774..1027  
 FT /\*tag= b  
 FT /\*number= 9  
 FT exon 1028..1124  
 FT /\*tag= c  
 FT /\*number= 10  
 XX  
 PN WO200179232-A2.  
 XX  
 XX 25-OCT-2001.  
 PD  
 PF 16-APR-2001; 2001WO-US12301.  
 PF  
 PR 14-APR-2000; 2000US-197839P.  
 PR  
 XX (GENA-) GENAISSANCE PHARM INC.  
 PA  
 XX Chew A, Choi JY, Koshy B, Stephens JC;  
 PI  
 XX WPI; 2002-055251/07.  
 DR  
 XX P-PSDB; AAU09495.  
 XX  
 PT Nucleotide polymorphisms in the human myeloproliferative leukemia virus

PT oncogene (MPL) gene, useful for studying the function of and expressing  
 PT MPL protein for use in screening drugs for treating diseases related to  
 PT MPL activity -  
 XX  
 XX Claim 27; Fig 3; 85pp; English.  
 XX  
 CC The present invention relates to novel single nucleotide polymorphisms  
 CC (SNPs) in the human myeloproliferative leukaemia virus oncogene (MPL)  
 CC gene located on chromosome 1p34, and methods for haplotyping and/or  
 CC genotyping the MPL gene. The methods of the invention make use of  
 CC allele-specific oligonucleotides (ASOs) as probes and primers and/or  
 CC primer-extension oligonucleotides for detecting MPL gene polymorphisms.  
 CC The polynucleotides and screened compounds are useful for the  
 CC treatment of diseases associated with MPL activity, such as  
 CC congenital amegakaryocytic thrombocytopaenia (CAMT). The present  
 CC sequence represents a reference sequence for human MPL gene exons 9-10.  
 CC Note: This sequence encodes for only part of the MPL protein shown  
 CC in AAU09495.  
 XX  
 SQ Sequence 1785 BP; 317 A; 562 C; 500 G; 403 T; 3 other;

Query Match 71.7%; Score 17.2; DB 24; Length 1785;  
 Best Local Similarity 86.4%; Pred. No. 3e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcagccagttccaccctgtca 22  
 ||||| ||||| ||||| |||||  
 Db 1672 tcagcgtgtctaccctgtgca 1693

RESULT 8  
 AAS91810/c  
 ID AAS91810 standard; cDNA; 1972 BP.  
 XX AC AAS91810;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE DNA encoding novel human diagnostic protein #27614.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 PF 30-MAR-2001; 2001WO-US08631.  
 PF  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 XX Drmanac RT, Liu C, Tang YT;  
 PI  
 XX WPI; 2001-639362/73.  
 DR  
 DR P-PSDB; ABG27623.  
 XX  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PT  
 XX Claim 1; SEQ ID No 27614; 103pp; English.  
 PS  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC the polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AA564197-AA594564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1972 BP; 460 A; 523 C; 552 G; 437 T; 0 other;

Query Match 71.7%; Score 17.2; DB 23; Length 1972;  
 Best Local Similarity 86.4%; Pred. No. 3.1e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 cagccaggtccaccctgac 23  
 ||||| ||||| ||||| |||||  
 DB 705 CCGCCAGTCCACACCCGTGCAC 684

RESULT 9  
 AAT90830  
 ID AAT90830 standard; cDNA; 1996 BP.  
 AC AAT90830;  
 XX  
 DT 15-JAN-1998 (first entry)  
 DE cDNA encoding glucoamylase P.  
 KW Glucoamylase P; hormoconis resinase; debranching activity; enzyme;  
 KW dextrinase activity; Trichoderma reesei; starch granule; preservation;  
 KW hydrolysis; gelatinised starch; glucose syrup; straight linear dextrin;  
 KW saccharification; lignocellulosic material; sugar utilisation;  
 KW feed additive; ss.  
 XX Hormoconis resinase.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 30..1880  
 FT /\*tag= a  
 FT polyA\_signal 2515..2520  
 FT /\*tag= b  
 XX  
 PN US5665585-A.  
 XX  
 XX 09-SEP-1997.  
 XX  
 PF 03-SEP-1992; 92US-0937789.  
 XX  
 PR 12-AUG-1993; 93US-0104853.  
 PR 03-SEP-1992; 92US-0937789.  
 PR 07-FEB-1995; 95US-0385370.  
 XX  
 PA (ALKO-) ALKO-YHTIOT OY.  
 XX  
 XX Aho S, Fagerstroem R, Joutsjoki V, Korhola M, Nevalainen H;  
 PI Torckeli H, Torckeli T, Vainio A;  
 XX  
 DR WPI; 1997-456802/42.  
 DR P-FSDB; AAW30155.  
 XX  
 PT Hormoconis resinase glucoamylase P gene construct - for transforming

PT Trichoderma to produce recombinant glucoamylase P  
 XX  
 PS Claim 2; Column 55-58; 6lpp; English.

CC AAT90830 and AAT90831 represent the cDNA and gene encoding the  
 CC Hormoconis resinase glucoamylase P, respectively. These sequences are used  
 CC in the composition of the invention and are capable of being processed by  
 CC a Trichoderma host cell. H. resinase glucoamylase P has higher  
 CC debranching and dextrinase activity than conventional  
 CC glucoamylase-pullulanase mixtures. T. reesei secretes enzymes that are  
 CC important for the degradation of complexes around and in starch granules.  
 CC The recombinant glucoamylase P enzyme produced by the Trichoderma host  
 CC cell is useful in applications requiring the hydrolysis of gelatinised  
 CC starch or the presence of a debranching activity. These applications  
 CC include starch analysis, the manufacture of glucose syrups, production of  
 CC straight linear dextrin for use in food, medicines and cosmetics, and in  
 CC the preparation of food fibers by the enzymatic treatment of seed husks  
 CC or brans. The enzyme can also be used as an additive to laundry and dish  
 CC washing detergents, in wood and textile industry applications such as  
 CC the preparation of plywood adhesives. It can also be used in the  
 CC saccharification of lignocellulosic materials, the preservation of  
 CC protein-containing animal or vegetable fodder, the preparation of  
 CC feedstuff from meat by-products and to improve utilisation of sugar  
 CC present in such feedstuff. When used as a feed additive for food  
 CC animals, the recombinant bacterial biomass may be added directly to the  
 CC animal's fodder.  
 XX  
 SQ Sequence 1996 BP; 474 A; 574 C; 491 G; 457 T; 0 other;

Query Match 71.7%; Score 17.2; DB 18; Length 1996;  
 Best Local Similarity 86.4%; Pred. No. 3.1e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 tcagccaggtccaccctgca 22  
 || ||||| ||||| ||||| |||||  
 DB 1473 tcgcagctccaccctggca 1494

RESULT 10  
 AAX61116/c  
 ID AAX61116 standard; DNA; 2300 BP.  
 XX  
 AC AAX61116;  
 XX  
 DT 27-JUL-1999 (first entry)  
 XX  
 DE Human beta2-adrenergic receptor gene.  
 XX  
 KW Alkalb-adrenergic receptor; human; cardiovascular disease;  
 KW beta2 adrenergic receptor; genetic variation identification; hypertrophy;  
 KW asthma diagnosis; hypertension; prostatic disease; pulmonary disorder;  
 KW asthma; peripheral vascular disorder; neuropsychic disorder;  
 KW endocrine-metabolic disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9924454-A1.  
 XX  
 PD 20-MAY-1999.  
 XX  
 PF 04-NOV-1998; 98WO-US23496.  
 PR 10-NOV-1997; 97US-0086232.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 XX Buescher R, Hertmann V, Insel PA;  
 XX WPI; 1999-327357/27.  
 XX  
 PT Pairs of oligonucleotides for amplifying adrenergic receptor genes  
 XX

PS Disclosure; Fig 2; 58pp; English.

XX This sequence represents the human beta2-adrenergic receptor gene, and  
CC is amplified by the primers of the invention. The primers are non-self  
CC hybridizing; contain at least 15 nucleotides (nt) and has a melting  
CC temperature 50-85 deg. C. Each pair of primers is: non-cross-hybridizing;  
CC anneals to two distinct segments (separated by at least 400 nt); and  
CC generates a homogeneous population of gene segments in a polymerase chain  
CC reaction (PCR). At least one primer in the pair can extend a 3'-end  
CC sequence complementary to a template sequence in a DNA polymerase  
CC reaction. The primers are used to amplify segments of the alpha1B and  
CC beta2 adrenergic receptor genes, particularly to identify genetic  
CC variations for diagnosis of disease. Specifically variations in the  
CC alpha1B gene are associated with cardiovascular disease, hypertension and  
CC prostatic disease (hypertrophy), and those in the beta2 gene with  
CC cardiovascular disease, hypertension and asthma, but variations may also  
CC be associated with peripheral vascular, pulmonary, neuropsychic and  
CC endocrine-metabolic disorders. These primers allow rapid and specific  
CC amplification of large and homogeneous gene segments of the alpha1B and  
CC beta2 genes from a complex mixture of DNAs. This makes possible detection  
CC of genetic alterations not previously amenable to routine, automated and  
CC large-scale sequencing analysis.

XX Sequence 2300 BP; 495 A; 613 C; 646 G; 546 T; 0 other;

Query Match 71.7%; Score 17.2; DB 20; Length 2300;  
Best Local Similarity 86.4%; Pred. No. 3.1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cagccagttccaccctgac 23  
| ||||| ||||| ||||| |  
Db 616 CTGCCAGTTCACGCCGTGTC 595

RESULT 11

AAA38340/C

ID AAA38340 standard; DNA; 2305 BP.

XX AAA38340;

XX 21-AUG-2000 (first entry)

XX Human beta-adrenergic receptor-2 coding region.

XX Beta-adrenergic receptor-2 gene; coding region;  
XX polymorphism; polymorphic marker; cardiovascular disease;  
XX myocardial infarction; unstable angina; hypertension; atherosclerosis;  
XX stroke; prognosis; drug screening; treatment outcome; human; ds.

XX Homo sapiens.

XX WO200022166-A2.

XX 20-APR-2000.

XX 13-OCT-1999; 99WO-IB01678.

XX 14-OCT-1998; 98US-0104286.

XX 14-OCT-1998; 98US-0104302.

XX (EURO-) EUROPA MEDICAL AB.

XX Norberg LT, Andersson MK, Lindstrom PHR, Jonsson L;

XX WPI; 2000-318010/27.

XX Assessing cardiovascular status in humans involves comparing test  
XX polymorphic pattern comprising polymorphic positions within genes  
XX encoding specific proteins, with reference polymorphic pattern -

XX Disclosure; Page 124-125; 126pp; English.

XX

CC The invention relates to a novel method of assessing the cardiovascular  
CC status in an individual and to newly identified polymorphisms in the  
CC genes encoding angiotensin-converting enzyme (ACE), angiotensin II  
CC receptor type 1 (AT1) and type 2 (AT2), angiotensinogen (AGT), renin,  
CC aldosterone synthase, endothelin receptor type A and beta-adrenergic  
CC receptors 1 and 2. The method comprises determining the sequence at one  
CC or more polymorphic positions within these genes, and comparing the  
CC pattern of polymorphisms from the individual with a reference polymorphic  
CC pattern obtained from a population of individuals exhibiting a  
CC predetermined cardiovascular disease status. The polymorphic markers are  
CC useful for determining the predisposition of an individual to  
CC cardiovascular disorders such as myocardial infarction, unstable angina,  
CC hypertension, atherosclerosis and stroke. They are also useful for  
CC predicting the likely cardiovascular status of a patient given a  
CC treatment regimen comprising administration of cardiovascular drugs  
CC (e.g., ACE inhibitors, beta-adrenergic receptor antagonists (beta-  
CC blockers) or calcium channel blockers). One or more polymorphic markers  
CC provides a basis for predicting the outcome of a treatment regimen.  
CC Fragments of the genes comprising a polymorphic site may be used as  
CC primers and probes for detecting genetic polymorphisms or in molecular  
CC library arrays for high throughput screening. The genes, and the proteins  
CC they encode are useful in the screening of potential cardiovascular  
CC drugs. Determination of an individual's polymorphic pattern reduces or  
CC eliminates trial and error in selecting a treatment for a particular  
CC individual cardiovascular patient. It also provides the ability to  
CC eliminate patients from clinical trials who are predicted to be  
CC non-responsive, or at a risk for an adverse response, to a particular  
CC treatment regimen. Adverse results in an early trial can be evaluated to  
CC identify polymorphic patterns so that the adverse results can be  
CC correlated with a sub-population of the test population, permitting  
CC exclusion of such sub-populations from the treatment group. Beneficial  
CC drugs can be approved for use in the appropriate population, thereby  
CC decreasing the number of patients required for a clinical trial, which in  
CC turn decreases the duration and cost of such trials. The present  
CC sequence represents the human beta-adrenergic receptor-2 gene  
CC coding region (GenBank Y00106/g293708). The polymorphic sites identified  
CC are 839A/G, 872C/G, 1045A/G, 1284C/T, 1316A/C, 1845C/G, 2032A/G,  
CC 2068 no insert/G/C and 2070 no insert/C.

XX Sequence 2305 BP; 495 A; 616 C; 649 G; 545 T; 0 other;

Query Match 71.7%; Score 17.2; DB 21; Length 2305;  
Best Local Similarity 86.4%; Pred. No. 3.1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cagccagttccaccctgac 23

| ||||| ||||| ||||| |  
Db 616 CTGCCAGTTCACGCCGTGTC 595

RESULT 12

AAA38784/C

ID AAA38784 standard; DNA; 2340 BP.

XX AAA38784;

XX 05-OCT-2000 (first entry)

XX Human beta2 adrenergic receptor beta2AR gene.

XX Human; adrenergic receptor; beta2 adrenergic receptor; beta2AR;  
XX chromosome 5q31(12); disease predisposition; asthma; hypertension;  
XX congestive heart failure; ischemic heart disease; arrhythmia;  
XX obesity; diabetes; vascular disease; premature labour; migraine;  
XX anaphylaxis; chronic obstructive pulmonary disease; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

XX 1487..2340

XX CDS /\*tag= a

XX /product= "beta2 adrenergic receptor"



FT /note= "no stop codon given at 3' end of sequence"

FT /partial

FT 1487..1546

FT sig\_peptide

FT /tag= b

FT /label= 5'\_leader\_cistron

FT replace(1541,T)

FT /tag= c

FT 1588..2340

FT mat\_peptide

FT /tag= d

XX WC200031307-A1.

XX

XX 02-JUN-2000.

XX

XX 24-NOV-1999; 99WO-US27963.

XX

XX 25-NOV-1998; 98US-0109886.

XX

XX (UYCI-) UNIV CINCINNATI.

XX

XX Liggett SB;

XX

XX WPI; 2000-400107/34.

XX

XX Polymorphisms in the leader cistron (LC) of the beta 2-adrenergic

XX receptor (beta 2 AR), useful for predicting genetic disposition to a

XX disease modified by beta 2 AR expression e.g. congestive heart failure,

XX hypertension -

XX

XX Disclosure; Figure 1; 56pp; English.

XX

XX The present sequence is a fragment of the C allele of the human beta2

XX adrenergic receptor (beta2AR) gene, which is located on chromosome

XX 5q31 (12). The gene has two different alleles, and it has been shown that

XX the presence of two copies of the r allele leads to higher expression of

XX the gene. This is because the polymorphism is found in the 5' leader

XX sequence, which encodes a peptide which regulates expression of the

XX beta2AR gene. The polymorphism is thought to affect individuals'

XX responses to beta-agonists and beta-antagonists, and is likely to

XX influence their predisposition to asthma, hypertension,

XX congestive heart failure, ischemic heart disease, arrhythmia, obesity,

XX diabetes, vascular disease, premature labour, migraine, anaphylaxis and

XX chronic obstructive pulmonary disease (COPD). The gene can, therefore, be

XX used to predict the susceptibility of an individual to these diseases and

XX determine the best treatment.

XX

XX Sequence 2340 BP; 498 A; 627 C; 653 G; 562 T; 0 other;

XX

Query Match 71.7%; Score 17.2; DB 21; Length 2340;

Best Local Similarity 86.4%; Pred. No. 3.1e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 cagccagttccaccocctgcac 23

Db 1410 CTGCCAGTTCAGCCCGCTGTC 1389

RESULT 13

AAT90831

ID AAT90831 standard; cDNA; 2745 BP.

XX

XX AAT90831;

XX

XX 15-JAN-1998 (first entry)

XX

XX Glucoamylase P gene.

XX

XX Glucoamylase P; hormoconis resinae; debranching activity; enzyme;

XX dextrinase activity; Trichoderma reesei; starch granule; preservation;

XX hydrolysis; gelatinised starch; glucose syrup; straight linear dextrin;

XX saccharification; lignocellulosic material; sugar utilisation;

XX feed additive; ss.

XX Hormoconis resinae.

OS

XX Key Location/Qualifiers

PH 5'UTR 1..294

FT /tag= a

FT CAAT\_signal 87..90

FT /tag= b

FT TATA\_signal 225..229

FT /tag= c

FT exon 295..520

FT /tag= d

FT /number= 1

FT /note= "encodes residues 1 to 75"

FT intron 521..593

FT /tag= e

FT /number= 1

FT /note= "designated IVS (intervening sequence) 1"

FT exon 594..898

FT /tag= f

FT /number= 2

FT /note= "encodes residues 76 to 177"

FT intron 899..1074

FT /tag= g

FT /number= 2

FT /note= "designated IVS (intervening sequence) 2"

FT exon 1075..1583

FT /tag= h

FT /number= 3

FT /note= "encodes residues 178 to 347"

FT intron 1584..1636

FT /tag= i

FT /number= 3

FT /note= "designated IVS (intervening sequence) 3"

FT exon 1637..2447

FT /tag= j

FT /number= 4

FT /note= "encodes residues 348 to 616"

FT 3'UTR 2448..2745

FT /tag= k

FT polyA\_signal 2515..2520

FT /tag= l

XX US5665585-A.

XX

XX 09-SEP-1997.

XX

XX 03-SEP-1992; 92US-0937789.

XX

XX 12-AUG-1993; 93US-0104853.

XX 03-SEP-1992; 92US-0937789.

XX 07-FEB-1995; 95US-0385370.

XX

XX (ALKO-) ALKO-YHTIOT OY.

PA

XX Aho S, Fagerstrom R, Joutsjoki V, Korhola M, Nevalainen H;

XX Torckeli H, Torckeli T, Vainio A;

XX

XX WPI; 1997-456802/42.

DR P-PSDB; AAW30155.

XX

XX Hormoconis resinae glucoamylase P gene construct - for transforming

XX Trichoderma to produce recombinant glucoamylase P

XX

XX Claim 2; Column 60-66; 61pp; English.

XX

XX AAT90830 and AAT90831 represent the cDNA and gene encoding the

XX Hormoconis resinae glucoamylase P, respectively. These sequences are used

XX in the composition of the invention and are capable of being processed by

XX a Trichoderma host cell. H. resinae glucoamylase P has higher

XX debranching and dextrinase activity than conventional

XX glucoamylase-pullulanase mixtures. T. reesei secretes enzymes that are

XX important for the degradation of complexes around and in starch granules.







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:01:55 ; Search time 6034.22 Seconds  
(without alignments)  
83.231 Million cell updates/sec

Title: US-09-899-718a-7

Perfect score: 24  
Sequence: 1 tcagccagttccaccgccgtgcag 24

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_ov.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
|------------|-------|-------------|--------|----|-------------|

|   |      |       |        |    |           |
|---|------|-------|--------|----|-----------|
| 1 | 24   | 100.0 | 24     | 6  | AX349069  |
| 2 | 24   | 100.0 | 3785   | 6  | AX349063  |
| 3 | 19.2 | 80.0  | 177696 | 2  | AC103298  |
| c | 4    | 18.8  | 78.3   | 33 | AC028268  |
|   | 5    | 18.8  | 78.3   | 33 | AC080286  |
|   | 6    | 18.8  | 78.3   | 33 | AC072500  |
| c | 7    | 18.8  | 78.3   | 33 | AC070455  |
|   | 8    | 18.8  | 78.3   | 33 | AC0591790 |
| c | 9    | 18.4  | 76.7   | 33 | AC089614  |
|   | 10   | 18.4  | 76.7   | 33 | AC101702  |
| c | 11   | 18.4  | 76.7   | 33 | AC099583  |
|   | 12   | 18.2  | 75.8   | 33 | AC074630  |
| c | 13   | 18.2  | 75.8   | 33 | AC062755  |
|   | 14   | 18.2  | 75.8   | 33 | AC074924  |
|   | 15   | 18.2  | 75.8   | 33 | AC074924  |
|   | 16   | 18.2  | 75.8   | 33 | AC074924  |
|   | 17   | 18.2  | 75.8   | 33 | AC074924  |
|   | 18   | 18.2  | 75.8   | 33 | AC074924  |
|   | 19   | 18.2  | 75.8   | 33 | AC074924  |
| c | 20   | 18.2  | 75.8   | 33 | AC074924  |
| c | 21   | 18.2  | 75.8   | 33 | AC074924  |
| c | 22   | 18.2  | 75.8   | 33 | AC074924  |
|   | 23   | 18.2  | 75.8   | 33 | AC074924  |
|   | 24   | 18.2  | 75.8   | 33 | AC074924  |
|   | 25   | 18.2  | 75.8   | 33 | AC074924  |
|   | 26   | 18.2  | 75.8   | 33 | AC074924  |
|   | 27   | 18.2  | 75.8   | 33 | AC074924  |
|   | 28   | 18.2  | 75.8   | 33 | AC074924  |
|   | 29   | 18.2  | 75.8   | 33 | AC074924  |
| c | 30   | 18.2  | 75.8   | 33 | AC074924  |
| c | 31   | 18.2  | 75.8   | 33 | AC074924  |
|   | 32   | 18.2  | 75.8   | 33 | AC074924  |
|   | 33   | 18.2  | 75.8   | 33 | AC074924  |
| c | 34   | 17.8  | 74.2   | 33 | AC074924  |
|   | 35   | 17.8  | 74.2   | 33 | AC074924  |
| c | 36   | 17.8  | 74.2   | 33 | AC074924  |
| c | 37   | 17.8  | 74.2   | 33 | AC074924  |
|   | 38   | 17.8  | 74.2   | 33 | AC074924  |
| c | 39   | 17.8  | 74.2   | 33 | AC074924  |
|   | 40   | 17.8  | 74.2   | 33 | AC074924  |
|   | 41   | 17.8  | 74.2   | 33 | AC074924  |
|   | 42   | 17.8  | 74.2   | 33 | AC074924  |
| c | 43   | 17.6  | 73.3   | 33 | AC074924  |
| c | 44   | 17.6  | 73.3   | 33 | AC074924  |
| c | 45   | 17.6  | 73.3   | 33 | AC074924  |

ALIGNMENTS

|            |                                                            |                      |                          |                          |       |        |                 |                 |
|------------|------------------------------------------------------------|----------------------|--------------------------|--------------------------|-------|--------|-----------------|-----------------|
| RESULT     | 1                                                          | AX349069             | Sequence                 | 7 from Patent WO0202785. | 24 bp | DNA    | linear          | PAT 06-FEB-2002 |
| LOCUS      | AX349069                                                   | Sequence             | 7 from Patent WO0202785. | 24 bp                    | DNA   | linear | PAT 06-FEB-2002 |                 |
| DEFINITION | AX349069                                                   | Sequence             | 7 from Patent WO0202785. | 24 bp                    | DNA   | linear | PAT 06-FEB-2002 |                 |
| ACCESSION  | AX349069                                                   | Sequence             | 7 from Patent WO0202785. | 24 bp                    | DNA   | linear | PAT 06-FEB-2002 |                 |
| VERSION    | AX349069.1                                                 | GI:18615104          |                          |                          |       |        |                 |                 |
| KEYWORDS   |                                                            | synthetic construct. |                          |                          |       |        |                 |                 |
| SOURCE     |                                                            | synthetic construct. |                          |                          |       |        |                 |                 |
| ORGANISM   |                                                            | artificial sequence. |                          |                          |       |        |                 |                 |
| REFERENCE  | 1 (sites)                                                  |                      |                          |                          |       |        |                 |                 |
| AUTHORS    | Sprunck,S., Kluth,A., Becker,D., Luetticke,S. and Loerz,H. |                      |                          |                          |       |        |                 |                 |
| TITLE      | Promoters of gene expression in plant caryopses            |                      |                          |                          |       |        |                 |                 |
| JOURNAL    | Patent: WO 0202785-A 7 10-JAN-2002;                        |                      |                          |                          |       |        |                 |                 |
| FEATURES   | Aventis CropScience GmbH (DE)                              |                      |                          |                          |       |        |                 |                 |
| source     | Location/Qualifiers                                        |                      |                          |                          |       |        |                 |                 |
|            | 1..24                                                      |                      |                          |                          |       |        |                 |                 |
|            | /organism="synthetic construct"                            |                      |                          |                          |       |        |                 |                 |
|            | /db_xref="taxon:32630"                                     |                      |                          |                          |       |        |                 |                 |
|            | /note="Oligonucleotide"                                    |                      |                          |                          |       |        |                 |                 |
|            | 4 a 11 c 5 g 4 t                                           |                      |                          |                          |       |        |                 |                 |

BASE COUNT 4 a 11 c 5 g 4 t  
ORIGIN

```

Query Match      100.0%; Score 24; DB 6; Length 24;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcagccagttccaccgcgtgcacg 24
    |||||||
DB 1 TCAGCCAGTTCCACCCTGTCACG 24

RESULT 2
LOCUS AX349063 3785 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 1 from Patent W00202785.
ACCESSION AX349063
VERSION AX349063.1 GI:18615098
KEYWORDS bread wheat.
SOURCE
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
REFERENCE 1 (sites)
AUTHORS Sprunck, S., Kluth, A., Becker, D., Luetticke, S. and Loez, H.
TITLE Promoters of gene expression in plant caryopses
JOURNAL Patent: WO 0202785-A 1 10-JAN-2002;
Aventis CropScience GmbH (DE)
FEATURES
source
1..3785
/organism="Triticum aestivum"
/db_xref="taxon:4565"
BASE COUNT 945 a 980 c 899 g 961 t
ORIGIN

Query Match      100.0%; Score 24; DB 6; Length 3785;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcagccagttccaccgcgtgcacg 24
    |||||||
DB 2927 TCAGCCAGTTCCACCCTGTCACG 2950

RESULT 3
LOCUS AC103298 177696 bp DNA linear HTG 21-DEC-2001
DEFINITION Rattus norvegicus clone CH230-75J23, *** SEQUENCING IN PROGRESS
ACCESSION AC103298
VERSION AC103298.2 GI:17974811
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 177696)
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amarantunge, H.C., Are, J.R., Banks, T., Barbara, J.,
Benton, J., Blin, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowling, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
Burke, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,

```

```

Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C.,
Hollins, B., Homsli, F., Howard, S., Huber, J., Huiyik, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Loulseged, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M.,
Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,
Ogih, M., Okuwon, G., Otagunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M.,
Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N.,
Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Thomas, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Washington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 177696)
Worley, K.C.
Direct Submission
Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 21, 2001 this sequence version replaced gi:17062960.
----- Genome Center of Medicine
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GJWU
Center clone name: CH230-75J23
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 142002 bases at least Q40
Consensus quality: 151283 bases at least Q30
Consensus quality: 159356 bases at least Q20
Estimated insert size: 148940; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.1x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 66 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 9784 9783: contig of 9783 bp in length
* 9884 9883: gap of unknown length
* 16894 16895: contig of 7012 bp in length
* 16896 16895: gap of unknown length
* 22488 22488: contig of 5493 bp in length
* 22489 22588: gap of unknown length
* 22589 27562: contig of 4974 bp in length
* 27563 27662: gap of unknown length
* 31812 31812: contig of 4150 bp in length
* 31912 31912: gap of unknown length
* 38604 38604: contig of 6692 bp in length
* 38704 38704: gap of unknown length

```

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 JOURNAL  
 COMMENT

```
* 38705 42885: contig of 4181 bp in length
* 42886 42985: gap of unknown length
* 42986 46890: contig of 3905 bp in length
* 46891 46990: gap of unknown length
* 46991 51589: contig of 4599 bp in length
* 51590 51689: gap of unknown length
* 51690 56276: contig of 4587 bp in length
* 56277 56376: gap of unknown length
* 56377 60645: contig of 4269 bp in length
* 60646 60745: gap of unknown length
* 60746 65606: contig of 4861 bp in length
* 65607 65706: gap of unknown length
* 65707 71186: contig of 5480 bp in length
* 71187 71286: gap of unknown length
* 71287 74378: contig of 3092 bp in length
* 74379 74478: gap of unknown length
* 74479 77718: contig of 3240 bp in length
* 77719 77818: gap of unknown length
* 77819 79827: contig of 1909 bp in length
* 79828 83044: gap of unknown length
* 83045 83144: contig of 3217 bp in length
* 83145 86172: contig of 3028 bp in length
* 86173 86272: gap of unknown length
* 86273 89912: contig of 3640 bp in length
* 89913 90012: gap of unknown length
* 90013 91949: contig of 1937 bp in length
* 91950 92049: gap of unknown length
* 92050 95086: contig of 3037 bp in length
* 95087 95186: gap of unknown length
* 95187 98089: contig of 2903 bp in length
* 98090 98189: gap of unknown length
* 98190 100105: contig of 1916 bp in length
* 100106 100205: gap of unknown length
* 100206 102487: contig of 2282 bp in length
* 102488 102587: gap of unknown length
* 102589 105472: contig of 2885 bp in length
* 105473 105572: gap of unknown length
* 105573 106955: contig of 1383 bp in length
* 106956 107055: gap of unknown length
* 107056 108993: contig of 1938 bp in length
* 108994 109093: gap of unknown length
* 109094 111378: contig of 2285 bp in length
* 111379 111478: gap of unknown length
* 111479 113502: contig of 2024 bp in length
* 113503 113602: gap of unknown length
* 113603 115962: contig of 2360 bp in length
* 115963 116062: gap of unknown length
* 116063 118662: contig of 2600 bp in length
* 118663 118762: gap of unknown length
* 118763 121876: contig of 3114 bp in length
* 121877 121976: gap of unknown length
* 121977 124696: contig of 2720 bp in length
* 124697 124796: gap of unknown length
* 124797 126707: contig of 1911 bp in length
* 126708 126807: gap of unknown length
* 126808 128988: contig of 2181 bp in length
* 128989 131431: contig of 2343 bp in length
* 131432 131433: gap of unknown length
* 131434 132655: contig of 1124 bp in length
* 132656 132755: gap of unknown length
* 132756 133997: contig of 1242 bp in length
* 133998 134097: gap of unknown length
* 134098 135946: contig of 1849 bp in length
* 135947 136046: gap of unknown length
* 136047 137778: contig of 1732 bp in length
* 137779 137878: gap of unknown length
* 137879 139111: contig of 1233 bp in length
* 139112 139211: gap of unknown length
* 139212 140450: contig of 1239 bp in length
* 140451 140550: gap of unknown length
* 140551 142742: contig of 2192 bp in length

* 142743 142842: gap of unknown length
* 142843 144888: contig of 2046 bp in length
* 144889 144988: gap of unknown length
* 144989 146161: contig of 1173 bp in length
* 146162 146261: gap of unknown length
* 146262 148193: contig of 1932 bp in length
* 148194 148293: gap of unknown length
* 148294 149569: contig of 1276 bp in length
* 149570 149669: gap of unknown length
* 149670 150978: contig of 1309 bp in length
* 150979 151078: gap of unknown length
* 151079 153090: contig of 2012 bp in length
* 153091 153190: gap of unknown length
* 153191 155093: contig of 1903 bp in length
* 155094 155193: gap of unknown length
* 155194 156320: contig of 1127 bp in length
* 156321 156420: gap of unknown length
* 156421 157945: contig of 1525 bp in length
* 157946 158045: gap of unknown length
* 158046 159462: contig of 1417 bp in length
* 159463 159562: gap of unknown length
* 159563 160575: contig of 1013 bp in length
* 160576 160675: gap of unknown length
* 160676 162165: contig of 1490 bp in length
* 162166 162265: gap of unknown length
* 162266 163643: contig of 1378 bp in length
* 163644 163743: gap of unknown length

Query Match 80.0%; Score 19.2; DB 2; Length 177696;
Best Local Similarity 87.5%; Pred. NO. 2.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcagccagttccaccctgtgacg 24
Db 91657 TCAGCCTCTTCCACCCTGTGCAG 91680

RESULT 4
AC028268/C standard; DNA; HTG; 959 BP.
ID AC028268
XX AC AC028268;
XX AC AC028268.1
SV AC028268.1
XX
XX
DT 05-APR-2000 (Rel. 63, Created)
DT 05-APR-2000 (Rel. 63, Last updated, Version 1)
XX
DE Giardia intestinalis clone CI0266 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.
XX
KW HTG; HTGS_PHASE0.
XX
OS Giardia intestinalis
OC Eukaryota; Diplomonadida; Hexamitidae; Giardia.
XX
XX [1]
RN 1-959
RP Morrison H.G., McArthur A.G., Nixon J., Eakin N.O., Kim U., Crocker M.C.,
RA Hinkle G., Holder M.E., Sogin M.L.;
RA "Giardia: a model for ancient eukaryotic genome analysis";
RT Unpublished.
XX
XX [2]
RN 1-959
RP Crocker M.K., Morrison H.G., McArthur A.G., Nixon J., Eakin N.O., Kim U.,
RA Hinkle G., Holder M.E., Sogin M.L.;
RT Submitted (01-APR-2000) to the EMBL/GenBank/DBJ databases.
RL Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
RL Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
XX
CC * NOTE: This record contains 1 individual
CC * sequencing reads that have not been assembled into
```

CC \* contigs. Runs of N are used to separate the reads  
 CC \* and the order in which they appear is completely  
 CC \* arbitrary. Low-pass sequence sampling is useful for  
 CC \* identifying clones that may be gene-rich and allows  
 CC \* overlap relationships among clones to be deduced.  
 CC \* However, it should not be assumed that this clone  
 CC \* will be sequenced to completion. In the event that  
 CC \* the record is updated, the accession number will  
 CC \* be preserved.  
 CC 1 959: contig of 959 bp in length.

XX Key Location/Qualifiers  
 FH source  
 FT 1.959  
 FT /db\_xref="taxon:5741"  
 FT /organism="Giardia intestinalis"  
 FT /strain="WB-C6"  
 FT /clone="CI0266"  
 XX Sequence 959 BP; 226 A; 251 C; 259 G; 216 T; 7 other;  
 SQ

Query Match 78.3%; Score 18.8; DB 33; Length 959;  
 Best Local Similarity 90.9%; Pred. No. 6.7e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 agccagttccaccccgtagc 24  
 ||||| || ||||| |||||  
 Db 265 AGCCAGGTCGACCCCGTGCACG 244

RESULT 5  
 AC080236 standard; DNA; HTG; 960 BP.  
 ID AC080286  
 AC AC080286;  
 XX AC080286.1  
 SV 23-SEP-2000 (Rel. 65, Created)  
 DT 29-SEP-2000 (Rel. 65, Last updated, Version 1)  
 DE Giardia intestinalis clone AJ2336 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.  
 XX HTG; HTGS\_PHASE0.  
 XX Giardia intestinalis  
 XX Eukaryota; Diplomonadida; Hexamitidae; Giardia.

[1]  
 RN Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,  
 RP Hinkle G., Holder M.E., Sogin M.L.;  
 RA "Giardia: a model for ancient eukaryotic genome analysis";  
 RT Unpublished.  
 RL  
 XX [2]  
 RN McArthur A.G., Morrison H.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,  
 RP Hinkle G., Holder M.E., Sogin M.L.;  
 RA Submitted (28-SEP-2000) to the EMBL/GenBank/DBJ databases.  
 RL Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,  
 RL Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA  
 XX

CC \* NOTE: This record contains 1 individual  
 CC \* sequencing reads that have not been assembled into  
 CC \* contigs. Runs of N are used to separate the reads  
 CC \* and the order in which they appear is completely  
 CC \* arbitrary. Low-pass sequence sampling is useful for  
 CC \* identifying clones that may be gene-rich and allows  
 CC \* overlap relationships among clones to be deduced.  
 CC \* However, it should not be assumed that this clone

CC \* will be sequenced to completion. In the event that  
 CC \* the record is updated, the accession number will  
 CC \* be preserved.  
 CC 1 960: contig of 960 bp in length.

XX Key Location/Qualifiers  
 FH source  
 FT 1.960  
 FT /db\_xref="taxon:5741"  
 FT /organism="Giardia intestinalis"  
 FT /strain="WB-C6"  
 FT /clone="AJ2336"  
 XX Sequence 960 BP; 219 A; 256 C; 248 G; 237 T; 0 other;  
 SQ

Query Match 78.3%; Score 18.8; DB 33; Length 960;  
 Best Local Similarity 90.9%; Pred. No. 6.7e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 agccagttccaccccgtagc 24  
 ||||| || ||||| |||||  
 Db 843 AGCCAGATCGACCCCGTGCACG 864

RESULT 6  
 AC072500 standard; DNA; HTG; 972 BP.  
 ID AC072500  
 AC AC072500;  
 XX AC072500.1  
 SV 13-JUN-2000 (Rel. 64, Created)  
 DT 13-JUN-2000 (Rel. 64, Last updated, Version 1)  
 DE Giardia intestinalis clone NU4295 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.  
 XX HTG; HTGS\_PHASE0.  
 XX Giardia intestinalis  
 XX Eukaryota; Diplomonadida; Hexamitidae; Giardia.

[1]  
 RN Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,  
 RP Hinkle G., Holder M.E., Sogin M.L.;  
 RA "Giardia: a model for ancient eukaryotic genome analysis";  
 RT Unpublished.  
 RL  
 XX [2]  
 RN Nixon J., Morrison H.G., McArthur A.G., Eakin N.Q., Kim U., Crocker M.C.,  
 RP Hinkle G., Holder M.E., Sogin M.L.;  
 RA Submitted (07-JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RL Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,  
 RL Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA  
 XX

CC \* NOTE: This record contains 1 individual  
 CC \* sequencing reads that have not been assembled into  
 CC \* contigs. Runs of N are used to separate the reads  
 CC \* and the order in which they appear is completely  
 CC \* arbitrary. Low-pass sequence sampling is useful for  
 CC \* identifying clones that may be gene-rich and allows  
 CC \* overlap relationships among clones to be deduced.  
 CC \* However, it should not be assumed that this clone  
 CC \* will be sequenced to completion. In the event that  
 CC \* the record is updated, the accession number will  
 CC \* be preserved.  
 CC 1 972: contig of 972 bp in length.

XX Key Location/Qualifiers  
 FH



|            |                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|------------|--------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| XX         | SQ                       | Sequence 990 BP; 228 A; 273 C; 282 G; 206 T; 1 other;                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|            | Query Match              | 78.3%; Score 18.8; DB 33; Length 990;                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|            | Best Local Similarity    | 90.9%; Pred. No. 6.7e+02;                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|            | Matches 20; Conservative | 0; Mismatches 2; Indels 0; Gaps 0;                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| QY         | 3                        | agccagtccaccgccgtgcacg 24<br>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Dd         | 800                      | AGCCAGATCGACCCGTCACG 779                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| RESULT     | 8                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| SME591790  |                          | 323450 bp DNA linear BCT 16-AUG-2001                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| LOCUS      |                          | Sinorhizobium meliloti 1021 complete chromosome; segment 9/12.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| DEFINITION |                          | AL591790 AL591688                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| VERSION    |                          | AL591790.1 GI:15075230                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| KEYWORDS   |                          | Sinorhizobium meliloti.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| SOURCE     |                          | Sinorhizobium meliloti.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| ORGANISM   |                          | Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Sinorhizobium.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| REFERENCE  |                          | Capela,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J., Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S., Godrie,F., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D., Pohl,T., Portetelle,D., Puehler,A., Purnelle,B., Ramsperger,U., Renard,C., Thebault,P., Vandenbol,M., Weidner,S. and Galibert,F.<br>From the Cover: Analysis of the chromosome sequence of the legume symbiont Sinorhizobium meliloti strain 1021                                                                                                                                                                                                                                                                                                                     |
| TITLE      |                          | Proceedings of the National Academy of Sciences of the United States of America. 98 (17), 9877-9882 (2001)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| JOURNAL    |                          | 11481430                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| PUBMED     |                          | 2 (bases 1 to 323450)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| REFERENCE  |                          | Gouzy,J.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| AUTHORS    |                          | Direct Submission                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| TITLE      |                          | Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO EU Consortium                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| JOURNAL    |                          | MELILO EU Consortium:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| COMMENT    |                          | Laboratoire de Biologie Molculaire des Relations Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet, France, Laboratoire de Genetique et Developpement UMR6061-CNRS, Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-33043 Rennes, France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany, Universitaet Bielefeld, Biologie IV (Genetik) Universitaetsstr 25, D-33615 Bielefeld, Germany, Unite de Biochimie physiologique, Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20, B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6, B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr http://sequence.toulouse.inra.fr/meliloti.html. |
| FEATURES   |                          | Location/Qualifiers<br>1. .323450<br>/organism="Sinorhizobium meliloti"<br>/strain="1021"<br>/db_xref="taxon:382"<br>155..1519<br>/gene="SMC01575"<br>155..1519<br>/gene="SMC01575"<br>/function="miscellaneous; hypothetical/partial homology"<br>/note="Product confidence : hypothetical<br>Gene name confidence : hypothetical<br>Predicted by Codon_usage<br>Predicted by Homology<br>Predicted by Framed"<br>/codon_start=1<br>/translation=1<br>/evidence=not_experimental<br>/product="HYPOTHETICAL PROTEIN"<br>/protein_id="CAC46787.1"                                                                                                                                                                                                                            |





```
* 3793 3892: gap of 100 bp
* 3893 4571: contig of 679 bp in length
* 4572 4671: gap of 100 bp
* 4672 5366: contig of 695 bp in length
* 5367 5466: gap of 100 bp
* 5467 6158: contig of 692 bp in length
* 6159 6258: gap of 100 bp
* 6259 6962: contig of 704 bp in length
* 6963 7062: gap of 100 bp
* 7063 7750: contig of 688 bp in length
* 7751 7850: gap of 100 bp
* 7851 8552: contig of 702 bp in length
* 8553 8652: gap of 100 bp
* 8653 9382: contig of 730 bp in length
* 9383 9482: gap of 100 bp
* 9483 10192: contig of 710 bp in length
* 10193 10292: gap of 100 bp
* 10293 10966: contig of 674 bp in length
* 10967 11066: gap of 100 bp
* 11067 11762: contig of 696 bp in length
* 11763 11862: gap of 100 bp
* 11863 12554: contig of 692 bp in length
* 12555 12654: gap of 100 bp
* 12655 13354: contig of 700 bp in length
* 13355 13454: gap of 100 bp
* 13455 14149: contig of 695 bp in length
* 14150 14249: gap of 100 bp
* 14250 14972: contig of 723 bp in length
* 14973 15072: gap of 100 bp
* 15073 15740: contig of 668 bp in length
* 15741 15840: gap of 100 bp
* 15841 16550: contig of 710 bp in length
* 16551 16650: gap of 100 bp
* 16651 17333: contig of 683 bp in length
* 17334 17433: gap of 100 bp
* 17434 18134: contig of 701 bp in length
* 18135 18234: gap of 100 bp
* 18235 18924: contig of 690 bp in length
* 18925 19024: gap of 100 bp
* 19025 19703: contig of 679 bp in length
* 19704 19803: gap of 100 bp
* 19804 20518: contig of 715 bp in length
* 20519 20618: gap of 100 bp
* 20619 21332: contig of 714 bp in length
* 21333 21432: gap of 100 bp
* 21433 22134: contig of 702 bp in length
* 22135 22234: gap of 100 bp
* 22235 22917: contig of 683 bp in length
* 22918 23017: gap of 100 bp
* 23018 23708: contig of 691 bp in length
* 23709 23808: gap of 100 bp
* 23809 24495: contig of 687 bp in length
* 24496 24595: gap of 100 bp
* 24596 25296: contig of 701 bp in length
* 25297 25396: gap of 100 bp
* 25397 26102: contig of 706 bp in length
* 26103 26202: gap of 100 bp
* 26203 26926: contig of 724 bp in length
* 26927 27026: gap of 100 bp
* 27027 27694: contig of 668 bp in length
* 27695 27794: gap of 100 bp
* 27795 28517: contig of 723 bp in length
* 28518 28617: gap of 100 bp
* 28618 29334: contig of 717 bp in length
* 29335 29434: gap of 100 bp
* 29435 30116: contig of 682 bp in length
* 30117 30216: gap of 100 bp
* 30217 30903: contig of 687 bp in length
* 30904 31003: gap of 100 bp
* 31004 31693: contig of 690 bp in length
* 31694 31793: gap of 100 bp
* 31794 32483: contig of 690 bp in length
* 32484 32583: gap of 100 bp
```

```
* 32584 33258: contig of 675 bp in length
* 33259 33358: gap of 100 bp
* 33359 34077: contig of 719 bp in length
* 34078 34177: gap of 100 bp
* 34178 34887: contig of 710 bp in length
* 34888 34987: gap of 100 bp
* 34988 35656: contig of 669 bp in length
* 35657 35756: gap of 100 bp
* 35757 36444: contig of 688 bp in length
* 36445 36544: gap of 100 bp
* 36545 37262: contig of 718 bp in length
* 37263 37362: gap of 100 bp
* 37363 38053: contig of 691 bp in length
* 38054 38153: gap of 100 bp
* 38154 38857: contig of 704 bp in length
* 38858 38957: gap of 100 bp
* 38958 39666: contig of 709 bp in length
* 39667 39766: gap of 100 bp
* 39767 40463: contig of 697 bp in length
* 40464 40563: gap of 100 bp
* 40564 41266: contig of 703 bp in length
* 41267 41366: gap of 100 bp
* 41367 42090: contig of 724 bp in length
* 42091 42190: gap of 100 bp
* 42191 42907: contig of 717 bp in length
* 42908 43007: gap of 100 bp
* 43008 43708: contig of 701 bp in length
* 43709 43808: gap of 100 bp
* 43809 44503: contig of 695 bp in length
* 44504 44603: gap of 100 bp
* 44604 45297: contig of 694 bp in length
* 45298 45397: gap of 100 bp
* 45398 46100: contig of 703 bp in length
* 46101 46200: gap of 100 bp
* 46201 46892: contig of 692 bp in length
* 46893 46992: gap of 100 bp
* 46993 47682: contig of 690 bp in length
* 47683 47782: gap of 100 bp
* 47783 48493: contig of 711 bp in length
* 48494 48593: gap of 100 bp
* 48594 49276: contig of 683 bp in length
* 49277 49376: gap of 100 bp
* 49377 50096: contig of 720 bp in length
* 50097 50196: gap of 100 bp
* 50197 50903: contig of 707 bp in length
* 50904 51003: gap of 100 bp
* 51004 51660: contig of 657 bp in length
* 51661 51760: gap of 100 bp
* 51761 52454: contig of 694 bp in length
* 52455 52554: gap of 100 bp
* 52555 53255: contig of 701 bp in length
* 53256 53355: gap of 100 bp
* 53356 54030: contig of 675 bp in length
* 54031 54130: gap of 100 bp
* 54131 54843: contig of 713 bp in length
* 54844 54943: gap of 100 bp
```

Query Match 76.7%; Score 18.4; DB 2; Length 113288;  
Best Local Similarity 95.0%; Pred. No. 6.3e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 ccagttccaccctgtgcacg 24  
|||||  
DB 99957 CCAGTTCACCCCGTGCAG 99938

## RESULT 11

AC099583

LOCUS

DEFINITION

unordered pieces.

AC099583

VERSION

AC099583 233684 bp DNA linear HTG 16-NOV-2001  
Mus musculus clone RP23-189G16, WORKING DRAFT SEQUENCE, 25  
AC099583  
GI:16946032

KEYWORDS  
SOURCE

HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

house mouse.

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 233684)

## AUTHORS

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 233684)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckghalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Menes, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trifilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (16-Nov-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A. F. A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## TITLE

## JOURNAL

## COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L16855

Center clone name: 189\_G16

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 225791 bases at least Q40

Consensus quality: 229156 bases at least Q30

Consensus quality: 230361 bases at least Q20

Insert size: 240000; agarose-fp

Quality size: 231284; sum-of-contigs

Quality coverage: 11.5 in Q20 bases; agarose-fp

Quality coverage: 11.9 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 25 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 458: contig of 458 bp in length

\* 459 558: gap of 100 bp

\* 559 2100: contig of 1542 bp in length

\* 2101 2200: gap of 100 bp

\* 2201 3332: contig of 1132 bp in length

\* 3333 3432: gap of 100 bp

\* 3433 4932: contig of 1560 bp in length

\* 4993 5092: gap of 100 bp

\* 5093 6602: contig of 1510 bp in length

\* 6603 6702: gap of 100 bp

\* 6703 10246: contig of 3544 bp in length

\* 10247 10346: gap of 100 bp

\* 10347 12129: contig of 1783 bp in length

\* 12130 12229: gap of 100 bp

\* 12230 14454: contig of 2225 bp in length

\* 14455 14554: gap of 100 bp

\* 14555 16558: contig of 2004 bp in length

\* 16559 16658: gap of 100 bp

\* 16659 19755: contig of 3097 bp in length

\* 19756 19855: gap of 100 bp

\* 19856 23363: contig of 3508 bp in length

\* 23364 23463: gap of 100 bp

\* 23464 27549: contig of 4085 bp in length

\* 27549 27648: gap of 100 bp

\* 27649 31362: contig of 3714 bp in length

\* 31363 31462: gap of 100 bp

\* 31463 36099: contig of 4637 bp in length

\* 36100 36199: gap of 100 bp

\* 36200 43798: contig of 7599 bp in length

\* 43799 43898: gap of 100 bp

\* 43899 51910: contig of 8012 bp in length

\* 51911 52010: gap of 100 bp

\* 52011 70202: contig of 18192 bp in length

\* 70203 70302: gap of 100 bp

\* 70303 102493: contig of 32191 bp in length

\* 102494 102593: gap of 100 bp

\* 102594 116654: contig of 14061 bp in length

\* 116655 116754: gap of 100 bp

\* 116755 129262: contig of 12508 bp in length

\* 129263 129362: gap of 100 bp

\* 129363 150868: contig of 21506 bp in length

\* 150869 150968: gap of 100 bp

\* 150969 176203: contig of 25235 bp in length

\* 176204 176303: gap of 100 bp

\* 176304 197144: contig of 20841 bp in length

\* 197145 197244: gap of 100 bp

\* 197245 221039: contig of 23795 bp in length

\* 221040 221339: gap of 100 bp

\* 221140 233684: contig of 12545 bp in length.

## FEATURES

## source

1. 233684

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="RP23-189G16"

/clone\_lib="RPCI-23 Female Mouse BAC"

## misc\_feature

1. 458

/note="assembly\_fragment"

clone\_end:SP6

vector\_side:left"

## misc\_feature

559..2100

/note="assembly\_fragment"

## misc\_feature

2201..3332

/note="assembly\_fragment"

## misc\_feature

3433..4992

/note="assembly\_fragment"

## misc\_feature

5093..6602

/note="assembly\_fragment"

## misc\_feature

6703..10246

/note="assembly\_fragment"

## misc\_feature

10347..12129

/note="assembly\_fragment"

## misc\_feature

12230..14454

/note="assembly\_fragment"

## misc\_feature

14555..16558

/note="assembly\_fragment"

## misc\_feature

16659..19755

/note="assembly\_fragment"

## misc\_feature

19856..23363

/note="assembly\_fragment"

## misc\_feature

23464..27549

```

misc_feature /note="assembly_fragment"
27649..31362
misc_feature /note="assembly_fragment"
31463..36099
misc_feature /note="assembly_fragment"
36200..43798
misc_feature /note="assembly_fragment"
43899..51910
misc_feature /note="assembly_fragment"
52011..70202
misc_feature /note="assembly_fragment"
70303..102493
misc_feature /note="assembly_fragment"
102594..116654
misc_feature /note="assembly_fragment"
116755..129262
misc_feature /note="assembly_fragment"
129363..150868
misc_feature /note="assembly_fragment"
150969..176203
misc_feature /note="assembly_fragment"
176304..197144
misc_feature /note="assembly_fragment"
197245..221039
misc_feature /note="assembly_fragment"
221140..233684
misc_feature /note="assembly_fragment"
clone_end:T7
vector_side:right"
BASE COUNT 65639 a 50317 c 51273 g 64053 t 2402 others
ORIGIN

Query Match 76.7%; Score 18.4; DB 2; Length 233684;
Best Local Similarity 95.0%; Pred. No. 5.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 cagccaggttcaccccggtgc 21
|||||
Db 210463 CAGCCAGTTCACCCGTCGTC 210482

RESULT 12
AC074630/c
AC AC074630 standard; DNA; HTG; 667 BP.
XX
AC AC074630;
XX
SV AC074630.1
XX
DT 02-AUG-2000 (Rel. 64, Created)
DT 02-AUG-2000 (Rel. 64, Last updated, Version 1)
XX
DE Giardia intestinalis clone AJ2486 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.
XX HTG; HTGS_PHASE0.
XX
OS Giardia intestinalis
OC Eukaryota; Diplomonadida; Hexamitidae; Giardia.
XX
RN [1]
RP Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
RA Hinkie G., Holder M.E., Sogin M.L.;
RT "Giardia: a model for ancient eukaryotic genome analysis";
RL Unpublished.
XX
RN [2]
RP McArthur A.G., Morrison H.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
RA Hinkie G., Holder M.E., Sogin M.L.;
RT ;
RL Submitted (31-JUL-2000) to the EMBL/GenBank/DBJ databases.

```

```

RL Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
RL Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
XX
CC * NOTE: This record contains 1 individual
CC * sequencing reads that have not been assembled into
CC * contigs. Runs of N are used to separate the reads
CC * and the order in which they appear is completely
CC * arbitrary. Low-pass sequence sampling is useful for
CC * identifying clones that may be gene-rich and allows
CC * overlap relationships among clones to be deduced.
CC * However, it should not be assumed that this clone
CC * will be sequenced to completion. In the event that
CC * the record is updated, the accession number will
CC * be preserved.
CC 1 667: contig of 667 bp in length.
XX Key Location/Qualifiers
FH source 1..667
FT /db_xref="taxon:5741"
FT /organism="Giardia intestinalis"
FT /strain="WB-C6"
FT /clone="AJ2486"
XX
SQ Sequence 667 BP; 159 A; 143 C; 183 G; 176 T; 6 other;

Query Match 75.8%; Score 18.2; DB 33; Length 667;
Best Local Similarity 87.0%; Pred. No. 1.3e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcagccaggttcaccccggtgcac 23
|||||
Db 661 TCAGCCAGTTCACCCGTCGTCAC 639

RESULT 13
HSA332832
LOCUS HSA332832 727 bp DNA linear PRI 01-OCT-2001
DEFINITION Homo sapiens genomic sequence surrounding NotI site, clone
HSJ-DG20RS.
ACCESSION AJ332832
VERSION AJ332832.1 GI:15877250
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 727)
AUTHORS Kutsenko, A.S., Gizatullin, R.Z., Al-Amin, A.N., Wang, F.,
Podowski, R.M., Matushkin, Y.G., Kvasha, S.M., Gyanchandani, A.,
Muravenko, O.V., Protopopov, A.I., Kashuba, V.I., Kisselev, L.L.,
Wasserman, W., Wahlstedt, C. and Zabarovsky, E.R.
TITLE Analysis of NotI flanking sequences: a new tool for gene discovery
and verification of the human genome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 727)
AUTHORS Zabarovsky, E.R.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-2001) Microbiology and Tumorigenology Centre,
Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
Sweden
FEATURES
source
1..727
/db_xref="taxon.9606"
/clone="HSJ-DG20RS"
BASE COUNT 166 a 239 c 217 g 104 t 1 others
ORIGIN

Query Match 75.8%; Score 18.2; DB 9; Length 727;
Best Local Similarity 87.0%; Pred. No. 1.3e+03;

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:23:47 ; Search time 165.21 Seconds  
(without alignments)  
17.842 Million cell updates/sec

Title: US-09-899-718A-6  
Perfect score: 12  
Sequence: 1 caggagcctcga 12

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, NA: \*  
1: /cgn2.6/prodata/2/ina/5A-COMB.seq: \*  
2: /cgn2.6/prodata/2/ina/5B-COMB.seq: \*  
3: /cgn2.6/prodata/2/ina/6A-COMB.seq: \*  
4: /cgn2.6/prodata/2/ina/6B-COMB.seq: \*  
5: /cgn2.6/prodata/2/ina/PCTUS-COMB.seq: \*  
6: /cgn2.6/prodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID                 | Description        |
|------------|-------|-------------|--------|--------------------|--------------------|
| c 1        | 12    | 100.0       | 475    | US-09-247-155-147  | Sequence 147, Appl |
| c 2        | 12    | 100.0       | 1341   | US-09-199-637A-266 | Sequence 266, Appl |
| c 3        | 12    | 100.0       | 1722   | US-09-221-275-3    | Sequence 3, Appl   |
| c 4        | 12    | 100.0       | 2001   | US-09-341-587-2    | Sequence 2, Appl   |
| c 5        | 12    | 100.0       | 2671   | US-09-221-275-2    | Sequence 2, Appl   |
| c 6        | 12    | 100.0       | 2678   | US-09-221-275-1    | Sequence 1, Appl   |
| c 7        | 12    | 100.0       | 3116   | US-08-449-643A-14  | Sequence 14, Appl  |
| c 8        | 12    | 100.0       | 3116   | US-08-702-367A-14  | Sequence 14, Appl  |
| c 9        | 12    | 100.0       | 3116   | PCT-US95-04681-14  | Sequence 14, Appl  |
| c 10       | 12    | 100.0       | 3348   | US-08-222-616-34   | Sequence 34, Appl  |
| c 11       | 12    | 100.0       | 3348   | PCT-US95-04228-34  | Sequence 34, Appl  |
| c 12       | 12    | 100.0       | 3995   | US-09-428-711A-13  | Sequence 13, Appl  |
| c 13       | 12    | 100.0       | 4680   | US-08-254-358-1    | Sequence 1, Appl   |
| c 14       | 12    | 100.0       | 4680   | US-08-475-391-1    | Sequence 1, Appl   |
| c 15       | 12    | 100.0       | 4680   | US-08-709-609-1    | Sequence 1, Appl   |
| c 16       | 12    | 100.0       | 4680   | PCT-US95-07178-1   | Sequence 1, Appl   |
| c 17       | 12    | 100.0       | 5802   | US-09-341-587-4    | Sequence 4, Appl   |
| c 18       | 12    | 100.0       | 28720  | US-09-341-587-7    | Sequence 7, Appl   |
| c 19       | 12    | 100.0       | 42235  | US-09-199-637A-1   | Sequence 1, Appl   |
| c 20       | 11    | 91.7        | 30     | US-08-836-504A-10  | Sequence 10, Appl  |
| c 21       | 11    | 91.7        | 97     | US-08-370-975B-9   | Sequence 9, Appl   |
| c 22       | 11    | 91.7        | 334    | US-08-358-171-22   | Sequence 22, Appl  |
| c 23       | 11    | 91.7        | 334    | US-09-090-947-22   | Sequence 22, Appl  |
| c 24       | 11    | 91.7        | 340    | US-08-594-031-83   | Sequence 83, Appl  |
| c 25       | 11    | 91.7        | 412    | US-08-047-041A-1   | Sequence 1, Appl   |
| c 26       | 11    | 91.7        | 412    | US-08-795-006A-21  | Sequence 21, Appl  |
| c 27       | 11    | 91.7        | 412    | US-09-184-073-21   | Sequence 21, Appl  |

|      |    |      |      |   |                   |                   |
|------|----|------|------|---|-------------------|-------------------|
| c 28 | 11 | 91.7 | 492  | 2 | US-08-892-880-12  | Sequence 12, Appl |
| c 29 | 11 | 91.7 | 528  | 1 | US-08-047-041A-14 | Sequence 14, Appl |
| c 30 | 11 | 91.7 | 534  | 1 | US-08-370-975B-13 | Sequence 13, Appl |
| c 31 | 11 | 91.7 | 624  | 1 | US-07-661-610C-14 | Sequence 14, Appl |
| c 32 | 11 | 91.7 | 642  | 1 | US-09-609-324A-9  | Sequence 9, Appl  |
| c 33 | 11 | 91.7 | 642  | 2 | US-08-920-440B-9  | Sequence 9, Appl  |
| c 34 | 11 | 91.7 | 642  | 4 | US-09-173-492-9   | Sequence 9, Appl  |
| c 35 | 11 | 91.7 | 642  | 4 | US-09-173-133-9   | Sequence 13, Appl |
| c 36 | 11 | 91.7 | 1025 | 4 | US-09-269-617-13  | Sequence 13, Appl |
| c 37 | 11 | 91.7 | 1329 | 1 | US-08-378-630A-8  | Sequence 8, Appl  |
| c 38 | 11 | 91.7 | 1454 | 4 | US-09-372-422A-19 | Sequence 19, Appl |
| c 39 | 11 | 91.7 | 1560 | 1 | US-08-356-180-1   | Sequence 1, Appl  |
| c 40 | 11 | 91.7 | 2313 | 2 | US-08-892-880-1   | Sequence 1, Appl  |
| c 41 | 11 | 91.7 | 2419 | 2 | US-08-765-662-13  | Sequence 13, Appl |
| c 42 | 11 | 91.7 | 2419 | 5 | PCT-US95-08745-13 | Sequence 13, Appl |
| c 43 | 11 | 91.7 | 2542 | 3 | US-08-941-445A-6  | Sequence 6, Appl  |
| c 44 | 11 | 91.7 | 3014 | 2 | US-08-808-982-1   | Sequence 1, Appl  |
| c 45 | 11 | 91.7 | 3014 | 4 | US-09-306-902A-1  | Sequence 1, Appl  |

## ALIGNMENTS

RESULT 1  
US-09-247-155-147/c  
; Sequence 147, Application US/09247155A  
; Patent No. 6312922  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste  
; APPLICANT: Duclert, Aymeric  
; APPLICANT: Bougueleret, Lydie  
; TITLE OF INVENTION: Complementary DNAs  
; FILE REFERENCE: GENSET.021A  
; CURRENT APPLICATION NUMBER: US/09/247,155A  
; CURRENT FILING DATE: 1999-02-09  
; EARLIER APPLICATION NUMBER: 60/074,121  
; EARLIER FILING DATE: 1998-02-09  
; EARLIER APPLICATION NUMBER: 60/081,563  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/096,116  
; EARLIER FILING DATE: 1998-08-10  
; EARLIER APPLICATION NUMBER: 60/099,273  
; EARLIER FILING DATE: 1998-10-04  
; NUMBER OF SEQ ID NOS: 182  
; SOFTWARE: Patent.pm  
; SEQ ID NO 147  
; LENGTH: 475  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 146..457  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 146..292  
; OTHER INFORMATION: Von Heijne matrix  
; OTHER INFORMATION: score 5.19999980926514  
; OTHER INFORMATION: seq CFCCLYPIPLCTS/HP  
; FEATURE:  
; NAME/KEY: polyA\_signal  
; LOCATION: 442..447  
; FEATURE:  
; NAME/KEY: polyA\_site  
; LOCATION: 465..475  
US-09-247-155-147

Query Match 100.0%; Score 12; DB 4; Length 475;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
|||||||

Db 223 CAGGAGCCTCGA 212

## RESULT 2

US-09-199-637A-266/c  
; Sequence 266, Application US/09199637A

; Patent No. 635411

; GENERAL INFORMATION:

; APPLICANT: Ausubel, Frederick

; APPLICANT: Goodman, Howard M.

; APPLICANT: Rahme, Laurence G.

; APPLICANT: Mahajan-Miklos, Shalina

; APPLICANT: Tan, Man-wah

; APPLICANT: Cao, Hui

; APPLICANT: Drenkard, Eliana

; APPLICANT: Tsongalis, John

; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID

; FILE REFERENCE: 00786/361002

; CURRENT APPLICATION NUMBER: US/09/199,637A

; PRIOR FILING DATE: 1998-11-25

; PRIOR FILING DATE: 1997-11-25

; NUMBER OF SEQ ID NOS: 437

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 266

; LENGTH: 1341

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-199-637A-266

## Query Match

Best Local Similarity 100.0%; Score 12; DB 4; Length 1341;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12

|||||

Db 669 CAGGAGCCTCGA 658

## RESULT 3

US-09-221-275-3

; Sequence 3, Application US/09221275

; Patent No. 6329332

; GENERAL INFORMATION:

; APPLICANT: Borneman, William S.

; APPLICANT: Bodie, Elizabeth A.

; APPLICANT: Graycar, Thomas P.

; APPLICANT: Winetzky, Deborah S.

; TITLE OF INVENTION: Pleurotus Phenol Oxidizing Enzymes

; FILE REFERENCE: GC557

; CURRENT APPLICATION NUMBER: US/09/221,275

; CURRENT FILING DATE: 1998-12-23

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 1722

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: cDNA

US-09-221-275-3

## Query Match

Best Local Similarity 100.0%; Score 12; DB 4; Length 1722;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12

|||||

Db 1230 caggagcctcga 1241

## RESULT 4

US-09-341-587-2/c

; Sequence 2, Application US/09341587

; Patent No. 6346606

; GENERAL INFORMATION:

; APPLICANT: Mollenhauer, Jan

; TITLE OF INVENTION: Protein Containing an SRCR Domain

; FILE REFERENCE: 4121-108

; CURRENT APPLICATION NUMBER: US/09/341,587

; CURRENT FILING DATE: 1999-08-31

; EARLIER APPLICATION NUMBER: PCT/DE98/00096

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 2001

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-341-587-2

## Query Match

Best Local Similarity 100.0%; Score 12; DB 4; Length 2001;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12

|||||

Db 248 CAGGAGCCTCGA 237

## RESULT 5

US-09-221-275-2

; Sequence 2, Application US/09221275

; Patent No. 6329332

; GENERAL INFORMATION:

; APPLICANT: Borneman, William S.

; APPLICANT: Bodie, Elizabeth A.

; APPLICANT: Graycar, Thomas P.

; APPLICANT: Winetzky, Deborah S.

; TITLE OF INVENTION: pleurotus Phenol Oxidizing Enzymes

; FILE REFERENCE: GC557

; CURRENT APPLICATION NUMBER: US/09/221,275

; CURRENT FILING DATE: 1998-12-23

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 2671

; TYPE: DNA

; ORGANISM: Pleurotus ostreatus

US-09-221-275-2

## Query Match

Best Local Similarity 100.0%; Score 12; DB 4; Length 2671;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12

|||||

Db 1976 caggagcctcga 1987

## RESULT 6

US-09-221-275-1

; Sequence 1, Application US/09221275

; Patent No. 6329332

; GENERAL INFORMATION:

; APPLICANT: Borneman, William S.

; APPLICANT: Bodie, Elizabeth A.

; APPLICANT: Graycar, Thomas P.

; APPLICANT: Winetzky, Deborah S.

; TITLE OF INVENTION: pleurotus Phenol Oxidizing Enzymes

; FILE REFERENCE: GC557

; CURRENT APPLICATION NUMBER: US/09/221,275  
; CURRENT FILING DATE: 1998-12-23  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 2678  
; TYPE: DNA  
; ORGANISM: Pleurotus ostreatus  
US-09-221-275-1

Query Match 100.0%; Score 12; DB 4; Length 2678;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
|||||

Db 1976 caggagcctcga 1987

RESULT 7  
US-08-449-645A-14/C  
; Sequence 14, Application US/08449645A  
; Patent No. 5981245  
; GENERAL INFORMATION:  
; APPLICANT: Fox, Gary M.  
; TITLE OF INVENTION: EPH-Like Receptor Protein Tyrosine  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Patent Operations/RBW  
; STREET: 1840 Behavilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/449,645A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Winter, Robert B.  
; REFERENCE/DOCKET NUMBER: A-287  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3116 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 34..2994  
US-08-449-645A-14

Query Match 100.0%; Score 12; DB 2; Length 3116;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
|||||

Db 731 CAGGAGCCTCGA 720

RESULT 8  
US-08-702-367A-14/C  
; Sequence 14, Application US/08702367A

; Patent No. 5981246  
; GENERAL INFORMATION:  
; APPLICANT: Fox, Gary M.  
; TITLE OF INVENTION: EPH-Like Receptor Protein Tyrosine  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Patent Operations/RBW  
; STREET: 1840 Behavilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/702,367A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Winter, Robert B.  
; REFERENCE/DOCKET NUMBER: A-287  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3116 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 34..2994  
US-08-702-367A-14

Query Match 100.0%; Score 12; DB 2; Length 3116;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
|||||

Db 731 CAGGAGCCTCGA 720

RESULT 9  
PCT-US95-04681-14/c  
; Sequence 14, Application PC/TUS9504681  
; GENERAL INFORMATION:  
; APPLICANT: Fox, Gary M.  
; TITLE OF INVENTION: EPH-Like Receptor Protein Tyrosine  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Patent Operations/RBW  
; STREET: 1840 Behavilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/04681  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Winter, Robert B.

REFERENCE/DOCKET NUMBER: A-287  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3116 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 34..2994  
PCT-US95-04681-14

Query Match 100.0%; Score 12; DB 5; Length 3116;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
|||||  
Db 731 CAGGAGCCTCGA 720

RESULT 10  
US-08-222-616-34/C  
Sequence 34, Application US/08222616  
Patent No. 5635177  
GENERAL INFORMATION:  
APPLICANT: Bennett, Brian D.  
APPLICANT: Goeddel, David  
APPLICANT: Lee, James M.  
APPLICANT: Matthews, William  
APPLICANT: Tsai, Siao Ping  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST ANTIBODIES  
TITLE OF INVENTION: ANTIBODIES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/222,616  
FILING DATE: 4-APR-1994  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/00586  
FILING DATE: 22-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/826935  
FILING DATE: 22-JAN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER:  
REFERENCE/DOCKET NUMBER: 821P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1994  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3348 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-222-616-34

Query Match 100.0%; Score 12; DB 1; Length 3348;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
|||||  
Db 698 CAGGAGCCTCGA 687

RESULT 11  
PCT-US95-04228-34/C  
Sequence 34, Application PC/TUS9504228  
GENERAL INFORMATION:  
APPLICANT: Genentech, Inc.  
APPLICANT: Bennett, Brian D.  
APPLICANT: Goeddel, David  
APPLICANT: Lee, James M.  
APPLICANT: Matthews, William  
APPLICANT: Tsai, Siao Ping  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST ANTIBODIES  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04228  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/222616  
FILING DATE: 04-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Wendy M. Lee  
REGISTRATION NUMBER: 00,000  
REFERENCE/DOCKET NUMBER: 821P3PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1994  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3348 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
PCT-US95-04228-34

Query Match 100.0%; Score 12; DB 5; Length 3348;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
|||||  
Db 698 CAGGAGCCTCGA 687

RESULT 12  
US-09-428-711A-13/C  
Sequence 13, Application US/09428711A

; Patent No. 6358720  
; GENERAL INFORMATION:  
; APPLICANT: Muramatsu, Masaaki  
; APPLICANT: Shirasawa, Takuji  
; APPLICANT: Tokumitsu, Hiroshi  
; APPLICANT: No. 6358720uchl, Teruhisa  
; TITLE OF INVENTION: SERINE/THREONINE PROTEIN KINASE  
; FILE REFERENCE: 06501-045001  
; CURRENT APPLICATION NUMBER: US/09/428,711A  
; CURRENT FILING DATE: 1999-10-28  
; PRIOR APPLICATION NUMBER: PCT/JP98/01246  
; PRIOR FILING DATE: 1998-03-23  
; PRIOR APPLICATION NUMBER: JP 9/124798  
; PRIOR FILING DATE: 1997-04-28  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 3995  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (388)...(3540)  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(3995)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-428-711A-13

Query Match 100.0%; Score 12; DB 4; Length 3995;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
|||||  
Db 2875 CAGGAGCCTCGA 2864

RESULT 13  
US-08-254-358-1  
; Sequence 1, Application US/08254358  
; Patent No. 5658785  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Philip R.  
; TITLE OF INVENTION: Adeno-Associated Virus Materials and  
; TITLE OF INVENTION: Methods  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 S. Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/254,358  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5658785and, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 31975  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 474-6300  
; TELEFAX: (312) 474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3995 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-254-358-1

; LENGTH: 4680 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-254-358-1

Query Match 100.0%; Score 12; DB 1; Length 4680;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
|||||  
Db 2993 CAGGAGCCTCGA 3004

RESULT 14  
US-08-475-391-1  
; Sequence 1, Application US/08475391  
; Patent No. 5786211  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Philip R.  
; TITLE OF INVENTION: Adeno-Associated Virus Materials and  
; TITLE OF INVENTION: Methods  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 S. Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/475,391  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/254,358  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5786211and, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 31975  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 474-6300  
; TELEFAX: (312) 474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4680 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-475-391-1

Query Match 100.0%; Score 12; DB 1; Length 4680;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
|||||  
Db 2993 CAGGAGCCTCGA 3004

RESULT 15

US-08-709-609-1  
; Sequence 1, Application US/08709609  
; Patent No. 5858775  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Phillip R.  
; TITLE OF INVENTION: Adeno-Associated Virus Materials and  
; TITLE OF INVENTION: Methods  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 S. Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08709,609  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5858775and, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 31975  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 474-6300  
; TELEFAX: (312) 474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4680 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-709-609-1

Query Match 100.0%; Score 12; DB 2; Length 4680;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 caggagcctcga 12  
| | | | | | | | | | | | | | | |  
Db 2993 CAGGAGCCTCGA 3004

Search completed: July 31, 2002, 12:23:49  
Job time: 11852 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:12:05 ; Search time 5855.71 Seconds  
(without alignments)  
27.659 Million cell updates/sec

Title: US-09-899-718A-6  
Perfect score: 12  
Sequence: 1 caqgaqcctcga 12

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:★

- ```

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estopl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Query % |       |        | DB  | ID       | Description        |                     |
|------------|---------|-------|--------|-----|----------|--------------------|---------------------|
|            | Score   | Match | Length |     |          |                    |                     |
| C 1        | 12      | 100.0 | 109    | 10  | BF371310 | BF371310 RC6-PN011 |                     |
|            | 2       | 12    | 100.0  | 110 | 12       | AZ747991           | AZ747991 RPCI-24-1  |
|            | 3       | 12    | 100.0  | 118 | 12       | BH354246           | BH354246 CH230-680  |
| C 4        | 4       | 12    | 100.0  | 129 | 12       | AZ748898           | AZ748898 RPCI-24-1  |
|            | 5       | 12    | 100.0  | 133 | 9        | AA584951           | AA584951 nn411b12.s |
|            | 6       | 12    | 100.0  | 144 | 10       | B1518243           | B1518243 603042090  |
| C 7        | 7       | 12    | 100.0  | 145 | 10       | BF812758           | BF812758 CM2-CI018  |
|            | 8       | 12    | 100.0  | 153 | 10       | BF926690           | BF926690 PM1-NT023  |
|            | 9       | 12    | 100.0  | 172 | 10       | BE827454           | BE827454 CM2-ET001  |
| C 10       | 10      | 12    | 100.0  | 173 | 10       | B1002833           | B1002833 MR3-HN016  |
|            | 11      | 12    | 100.0  | 176 | 12       | AZ271167           | AZ271167 RPCI-23-4  |
|            | 12      | 12    | 100.0  | 177 | 9        | AV640236           | AV640236 AV640236   |
| C 13       | 13      | 12    | 100.0  | 186 | 9        | AA038245           | AA038245 m182e03.r  |
|            | 14      | 12    | 100.0  | 193 | 9        | AA074114           | AA074114 z179f05.f  |
|            | 15      | 12    | 100.0  | 198 | 9        | AA657774           | AA657774 nu06f04.s  |
| C 16       | 16      | 12    | 100.0  | 198 | 9        | BB592607           | BB592607 BB592607   |
|            | 17      | 12    | 100.0  | 211 | 9        | AA770456           | AA770456 ah89f12.s  |

|    |    |       |       |     |          |                    |
|----|----|-------|-------|-----|----------|--------------------|
| 18 | 12 | 100.0 | 211   | 9   | AA450935 | AA450935 vf85b01.r |
| c  | 19 | 12    | 100.0 | 212 | 9        | BB588724           |
| c  | 20 | 12    | 100.0 | 212 | 9        | BE091664           |
| c  | 21 | 12    | 100.0 | 214 | 10       | BF505617           |
| c  | 22 | 12    | 100.0 | 215 | 10       | BM054775           |
| c  | 23 | 12    | 100.0 | 219 | 10       | BG275999           |
| c  | 24 | 12    | 100.0 | 229 | 9        | BB354103           |
| c  | 25 | 12    | 100.0 | 229 | 10       | BI051362           |
| c  | 26 | 12    | 100.0 | 229 | 10       | D21628             |
| c  | 27 | 12    | 100.0 | 230 | 10       | BE838749           |
| c  | 28 | 12    | 100.0 | 231 | 10       | BF402808           |
| c  | 29 | 12    | 100.0 | 232 | 10       | BF399261           |
| c  | 30 | 12    | 100.0 | 233 | 12       | BH204013           |
| c  | 31 | 12    | 100.0 | 249 | 10       | BM068805           |
| c  | 32 | 12    | 100.0 | 252 | 9        | AI020086           |
| c  | 33 | 12    | 100.0 | 253 | 10       | BE714112           |
| c  | 34 | 12    | 100.0 | 255 | 9        | AI337414           |
| c  | 35 | 12    | 100.0 | 258 | 10       | BF743579           |
| c  | 36 | 12    | 100.0 | 259 | 9        | AA365282           |
| c  | 37 | 12    | 100.0 | 260 | 9        | AV735802           |
| c  | 38 | 12    | 100.0 | 263 | 9        | BE143676           |
| c  | 39 | 12    | 100.0 | 265 | 10       | BI008774           |
| c  | 40 | 12    | 100.0 | 270 | 9        | AA348160           |
| c  | 41 | 12    | 100.0 | 270 | 9        | AA365578           |
| c  | 42 | 12    | 100.0 | 270 | 9        | BB590276           |
| c  | 43 | 12    | 100.0 | 271 | 9        | AI564849           |
| c  | 44 | 12    | 100.0 | 271 | 9        | AW325543           |
| c  | 45 | 12    | 100.0 | 272 | 10       | R22443             |
|    |    |       |       |     |          | R22443_vh23q06.s1  |

## ALIGNMENTS

RESULT 1  
BF371310/C

BF371310 109 bp mRNA linear EST 24-NOV-2000  
RC6-FN0114-090800-011-E11 FN0114 Homo sapiens cdNA, mRNA sequence.  
BF371310  
BF371310.1 GI:11333244  
EST.  
human.

ORGANISM

homo sapiens  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 109)  
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
 Nagai,M.A., da Silva,W. Jr., Zago,M.F., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A. J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

PROC. NA.  
20202663

COMMENT  
Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

**Tel: +55-11-2704922**

Fax: +55-11-2707001

Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)

This sequence was derived from the FAFESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?ti=RC66t2=RC6-FNO114-090800-011-Ell1&t3=2000-08-09&t4=1>)

Seq primer: puc 18 forward

High quality sequence start: 35

High quality sequence stop: 65.

Location/Qualifiers

1. .109

/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone_lib="FN0114"
/notes="Organ: prostate_normal; Vector: puc18; Site_1: SmaI
; Site_2: SmaI; A mini-library was made by cloning
products derived from ORSTES PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the pUC 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
BASE COUNT      27 a      33 c      26 g      23 t
ORIGIN

```

```

Query Match      100.0%; Score 12; DB 10; Length 109;
Best Local Similarity 100.0%; Pred. No. 8.6e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagctcga 12
|||||
Db 86 CAGGAGCCTCGA 75

```

```

RESULT 2
A2747991
LOCUS      A2747991      110 bp      DNA      linear      GSS 25-JAN-2001
DEFINITION RPCI-24-125D2.TV RPCI-24 Mus musculus genomic clone RPCI-24-125D2,
DNA sequence.
ACCESSION  A2747991
VERSION    A2747991.1 GI:12532400
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 110)
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Kroil,M., Shvartsbeyn,A., Gebregorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Other_GSSs: RPCI-24-125D2.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 125 row: D column: 2
Seq primer: T7
Class: BAC ends.

```

```

FEATURES
source
1. .110
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-125D2"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/notes="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT      34 a      24 c      25 g      27 t
ORIGIN

```

```

Query Match      100.0%; Score 12; DB 12; Length 110;
Best Local Similarity 100.0%; Pred. No. 8.7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagctcga 12
|||||
Db 27 CAGGAGCCTCGA 38

```

```

RESULT 3
BH354246/c
LOCUS      BH354246      118 bp      DNA      linear      GSS 03-DEC-2001
DEFINITION CH230-68019.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
CH230-68019, DNA sequence.
ACCESSION  BH354246
VERSION    BH354246.1 GI:17284980
KEYWORDS   GSS.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 118)
Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
Jong,P. and Fraser,C.M.
Rat BAC End Sequences from Library CHORI-230 EcoRI segment
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.choi.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.choi.org/bacpac/orering_informtion.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 68 row: O column: 19
Seq primer: T7
Class: BAC ends.

```

```

FEATURES
source
1. .118
/organism="Rattus norvegicus"
/strain="BN/SSNHsd/MCw"
/db_xref="taxon:10116"
/clone="CH230-68019"
/clone_lib="CHORI-230 Segment 1"
/sex="Female"
/cell_type="Brain"
/notes="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
CHORI-230 Rat (BN/SSNHsd/MCw) BAC library produced by
Pieter de Jong"
BASE COUNT      23 a      28 c      33 g      34 t
ORIGIN

```

```

Query Match      100.0%; Score 12; DB 12; Length 118;
Best Local Similarity 100.0%; Pred. No. 8.8e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagctcga 12
|||||
Db 33 CAGGAGCCTCGA 22

```

```

RESULT 4
AZ748998

```



LOCUS AZ748898 129 bp DNA linear GSS 25-JAN-2001  
 DEFINITION RPCI-24-125F2.TV RPCI-24 Mus musculus genomic clone RPCI-24-125F2,  
 DNA sequence.  
 ACCESSION AZ748898  
 VERSION AZ748898.1 GI:12534057  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 129)  
 Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Akinret, B., Lewins, M.,  
 Russell, D., de Jong, P. and Fraser, C.M.  
 Mouse BAC End Sequences from Library RPCI-24  
 Unpublished (1999)  
 JOURNAL Other\_GSSs: RPCI-24-125F2.TJ  
 COMMENT Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-24. For BAC  
 library availability, please contact Pieter de Jong  
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC  
 Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end  
 page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html  
 Plate: 125 row: F column: 2  
 Seq primer: T7  
 Class: BAC ends.

FEATURES  
 source Location/Qualifiers  
 1..129  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-24-125F2"  
 /clone\_lib="RPCI-24"  
 /sex="Male"  
 /cell\_type="Spleen/Brain"  
 /note="Vector: pTARBAC1; Site\_1: BamHI; Site\_2: BamHI;  
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The  
 library was cloned in the pTARBAC1 cloning vector at the  
 BamHI sites using MboI partially digested male C57BL/6J  
 DNA."  
 BASE COUNT 38 a 29 c 29 g 33 t  
 ORIGIN

Query Match 100.0%; Score 12; DB 12; Length 129;  
 Best Local Similarity 100.0%; Pred. No. 9e+03;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12  
 |||||  
 Db 28 CAGGAGCCTCGA 39

RESULT 5  
 AA584951  
 LOCUS  
 DEFINITION AA584951.1 NCI\_CGAP\_GC5 Homo sapiens cDNA clone IMAGE:1086431 3',  
 similar to SW:DHCA\_MOUSE P48758 CARBONYL REDUCTASE ;, mRNA  
 sequence.  
 ACCESSION AA584951  
 VERSION AA584951.1 GI:2367731  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 133)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: Stratagene, Inc.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality  
 Insert Length: 1082 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 1.  
 FEATURES Location/Qualifiers  
 1..133  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1086431"  
 /clone\_lib="NCI\_CGAP\_GC5"  
 /tissue\_type="germ cell tumor"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="Vector: Bluescript SK-; Site\_1: EcoRI; Site\_2: XhoI  
 cell tumors. 5' adaptor sequence: 5' GAATTCGGCAGG 3' 3',  
 adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' Average  
 insert size: 0.7 kb."  
 BASE COUNT 26 a 47 c 33 g 27 t  
 ORIGIN

Query Match 100.0%; Score 12; DB 9; Length 133;  
 Best Local Similarity 100.0%; Pred. No. 9.1e+03;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12  
 |||||  
 Db 86 CAGGAGCCTCGA 97

RESULT 6  
 BI518243  
 LOCUS  
 DEFINITION BI518243.1 NIH\_MGC\_116 Homo sapiens cdna clone IMAGE:5182415 3',  
 mRNA sequence.  
 ACCESSION BI518243  
 VERSION BI518243.1 GI:15343035  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 144)  
 NIH-MGC http://mgc.ncbi.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM11455 row: c column: 24  
 High quality sequence start: 24  
 High quality sequence stop: 144.

```

FEATURES
source
Location/Qualifiers
1. .144
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5182415"
/clone_lib="NIH_MGC_116"
/lab_host="DH10B"
/note="Organ: pooled colon, kidney, stomach; Vector:
pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon; 46 yo male kidney, and pool of 2
stomachs, 62 yo male and 70 yo female. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
023. Note: this is a NIH_MGC Library."
BASE COUNT      31 a  36 c  33 g  44 t
ORIGIN

Query Match      100.0%; Score 12; DB 10; Length 144;
Best Local Similarity 100.0%; Pred. No. 9.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||
Db 127 CAGCAGCCTCGA 138

RESULT 7
BF812758
LOCUS
DEFINITION
CM2-CI0180-221100-540-b07 CI0180 Homo sapiens cDNA, mRNA sequence.
BF812758
VERSION
BF812758.1 GI:12142691
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 145)
AUTHORS
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&t2=CM2-CI0180-
221100-540-b07&t3=2000-11-22&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 24
High quality sequence stop: 144.
Location/Qualifiers
1. .145
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CI0180"

FEATURES
source
Location/Qualifiers
1. .144
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5182415"
/clone_lib="NIH_MGC_116"
/lab_host="DH10B"
/note="Organ: pooled colon, kidney, stomach; Vector:
pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon; 46 yo male kidney, and pool of 2
stomachs, 62 yo male and 70 yo female. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
023. Note: this is a NIH_MGC Library."
BASE COUNT      31 a  36 c  33 g  44 t
ORIGIN

Query Match      100.0%; Score 12; DB 10; Length 144;
Best Local Similarity 100.0%; Pred. No. 9.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||
Db 94 CAGCAGCCTCGA 105

RESULT 8
BF926690/c
LOCUS
DEFINITION
PM1-NT0235-011200-002-a10 NT0235 Homo sapiens cDNA, mRNA sequence.
BF926690
ACCESSION
BF926690.1 GI:12324010
VERSION
EST.
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 153)
AUTHORS
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM1&t2=PM1-NT0235-
011200-002-a10&t3=2000-12-01&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 13
High quality sequence stop: 152.
Location/Qualifiers
1. .153
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NT0235"
/dev_stage="Adult"
/note="Organ: nervous tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      32 a  40 c  50 g  31 t
ORIGIN

/dev_stage="Adult"
/note="Organ: colon.ins; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      31 a  35 c  36 g  43 t
ORIGIN

Query Match      100.0%; Score 12; DB 10; Length 145;
Best Local Similarity 100.0%; Pred. No. 9.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||
Db 94 CAGCAGCCTCGA 105

```

Query Match 100.0%; Score 12; DB 10; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 9.5e+03;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
 |||||  
 Db 144 CAGGAGCCTCGA 133

RESULT 9  
 BE827454  
 LOCUS BE827454 172 bp mRNA linear EST 22-SEP-2000  
 DEFINITION CM2-ET0016-310500-206-d05 ET0016 Homo sapiens cDNA, mRNA sequence.  
 BE827454  
 ACCESSION BE827454.1 GI:10259819  
 VERSION BE827454.1  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 172)  
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags  
 PROC. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil

Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-ET0016-310500-206-d05et3=2000-05-31et4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 6  
 High quality sequence stop: 145.

FEATURES  
 source  
 1. .172  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="ET0016"  
 /dev\_stage="Adult"  
 /note="Organ: lung\_tumor; Vector: puc18; Site\_1: SmaI;  
 Site\_2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the puc 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."  
 BASE COUNT 35 a 49 c 35 g 53 t  
 ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 172;  
 Best Local Similarity 100.0%; Pred. No. 9.8e+03;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12  
 |||||  
 Db 55 CAGGAGCCTCGA 66

RESULT 10

BI002833 173 bp mRNA linear EST 13-JUN-2001  
 LOCUS MR3-HN0166-310101-002-e06 HN0166 Homo sapiens cDNA, mRNA sequence.  
 DEFINITION BI002833  
 ACCESSION BI002833.1 GI:14406907  
 VERSION BI002833.1  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 173)  
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 PROC. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil

Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=mr3=et2=MR3-HN0166-310101-002-e06et3=2001-01-31et4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 5  
 High quality sequence stop: 173.

FEATURES  
 Location/Qualifiers  
 1. 173  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="HN0166"  
 /dev\_stage="Adult"  
 /note="Organ: head\_normal; Vector: puc18; Site\_1: SmaI;  
 Site\_2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the puc 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."  
 BASE COUNT 36 a 59 c 50 g 28 t  
 ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 173;  
 Best Local Similarity 100.0%; Pred. No. 9.8e+03;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
 |||||  
 Db 162 CAGGAGCCTCGA 173

RESULT 11

AZ271167  
 LOCUS RPCI-23-444J17.TJ RPCI-23 Mus musculus genomic clone RPCI-23-444J17  
 DEFINITION AZ271167  
 , DNA sequence.  
 ACCESSION AZ271167  
 VERSION AZ271167.1 GI:9484784  
 KEYWORDS GSS.  
 SOURCE house mouse.

```

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 176)
AUTHORS Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akimret
, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(plate@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
plate: 444 row: J column: 17
Seq primer: SP6
Class: BAC ends.
FEATURES
    source
        1..176
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="RPCI-23-44A117"
            /clone_lib="RPCI-23"
            /sex="Female"
            /lab_host="DH10B"
            /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
            EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
            brain genomic DNA was isolated and partially digested
            with a combination of EcoRI and EcoRI Methylase. Size
            selected DNA was cloned into the pBACe3.6 vector at the
            EcoRI sites. The ligation products were transformed into
            DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 44 a 48 c 52 g 32 t
ORIGIN
    1 caggagcctcga 12
    |||||
    Db 12 CAGGAGCCTCGA 23

Query Match 100.0%; Score 12; DB 12; Length 176;
Best Local Similarity 100.0%; Pred. No. 9.8e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||
Db 12 CAGGAGCCTCGA 23

RESULT 12
AV640236 177 bp mRNA linear EST 15-DEC-2000
LOCUS AV640236 Chlamydomonas reinhardtii 5% CO2
DEFINITION cDNA clone HCL012fil_1 5', mRNA sequence.
ACCESSION AV640236
VERSION AV640236.1 GI:10783564
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 177)
AUTHORS Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,
Nakamura, Y., and Tabata, S.
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL DNA Res. 7 (5), 305-307 (2000)
MEDLINE 20539644

Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
    source
        1..177
            /organism="Chlamydomonas reinhardtii"
            /strain="C9"
            /db_xref="taxon:3055"
            /clone="HCL012fil_1"
            /clone_lib="Chlamydomonas reinhardtii 5% CO2"
            /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
            XhoI; The cDNA library was constructed from cells cultured
            in a medium with bubbling air containing 5% carbon
            dioxide"
BASE COUNT 34 a 56 c 57 g 30 t
ORIGIN
    1 caggagcctcga 12
    |||||
    Db 135 CAGGAGCCTCGA 146

Query Match 100.0%; Score 12; DB 9; Length 177;
Best Local Similarity 100.0%; Pred. No. 9.8e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||
Db 135 CAGGAGCCTCGA 146

RESULT 13
AA038245/c 186 bp mRNA linear EST 28-AUG-1996
LOCUS AA038245 Soares mouse p3NMFI9.5 Mus musculus cDNA clone
DEFINITION IMAGE:473116 5', mRNA sequence.
ACCESSION AA038245
VERSION AA038245.1 GI:1513669
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 186)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Willson, R. and
Waterston, R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:283860
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 180.
FEATURES
    source
        1..186
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="IMAGE:473116"
            /clone_lib="Soares mouse p3NMFI9.5"
            /dev_stage="19.5 dpc total fetus"
            /lab_host="DH10B (ampicillin resistant)"
            /note="Vector: pT73D (Pharmacia) with a modified
            polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
            was primed with a Not I - oligo(dT) primer [5'
            TGTTACCAATCTGAGTGGGAGCGCGCATTTTTTTTTTTT 3'],

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:11:22 ; Search time 720.8 seconds  
(without alignments)  
28.583 Million cell updates/sec

Title: US-09-899-718A-6  
Perfect score: 12  
Sequence: 1 caggagactcga 12

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues  
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| c 1        | 12    | 100.0       | 32     | 18 | PCR primer used in |
| c 2        | 12    | 100.0       | 58     | 22 | Oligonucleotide #2 |
| c 3        | 12    | 100.0       | 59     | 22 | Human albumin (HA) |
| c 4        | 12    | 100.0       | 59     | 22 | Human albumin (HA) |
| c 5        | 12    | 100.0       | 59     | 22 | Human albumin (HA) |
| c 6        | 12    | 100.0       | 193    | 22 | Human bone marrow  |
| c 7        | 12    | 100.0       | 261    | 23 | Drosophila melanog |
| c 8        | 12    | 100.0       | 292    | 21 | Human secreted pro |
| c 9        | 12    | 100.0       | 315    | 21 | Eucalyptus grandis |

|      |    |       |      |    |           |                    |
|------|----|-------|------|----|-----------|--------------------|
| c 10 | 12 | 100.0 | 341  | 21 | AACT75741 | Human ORFX ORF1296 |
| c 11 | 12 | 100.0 | 394  | 21 | AAC04396  | Human secreted pro |
| c 12 | 12 | 100.0 | 400  | 21 | AAC08874  | Human secreted pro |
| c 13 | 12 | 100.0 | 450  | 18 | AAI74033  | Aloe arborescens p |
| c 14 | 12 | 100.0 | 475  | 20 | AAZ40844  | Secreted protein E |
| c 15 | 12 | 100.0 | 552  | 22 | AAH97774  | Murine 7-transmemb |
| c 16 | 12 | 100.0 | 559  | 22 | AAH97770  | Murine 7-transmemb |
| c 17 | 12 | 100.0 | 568  | 22 | AAH10794  | Human cDNA clone ( |
| c 18 | 12 | 100.0 | 582  | 22 | AAK38542  | Human bone marrow  |
| c 19 | 12 | 100.0 | 616  | 21 | AAA43755  | Human secreted exp |
| c 20 | 12 | 100.0 | 633  | 22 | AAH97773  | Murine 7-transmemb |
| c 21 | 12 | 100.0 | 647  | 22 | AAH97771  | Murine 7-transmemb |
| c 22 | 12 | 100.0 | 652  | 22 | AAH97769  | Murine 7-transmemb |
| c 23 | 12 | 100.0 | 697  | 21 | AAC99009  | Human pancreatic c |
| c 24 | 12 | 100.0 | 822  | 19 | AAV28653  | Ripening banana pu |
| c 25 | 12 | 100.0 | 1271 | 21 | AAA69541  | Eucalyptus grandis |
| c 26 | 12 | 100.0 | 1341 | 20 | AAH98114  | Nucleotide sequenc |
| c 27 | 12 | 100.0 | 1527 | 21 | AAC33226  | Arabidopsis thalia |
| c 28 | 12 | 100.0 | 1563 | 23 | ABL23803  | Drosophila melanog |
| c 29 | 12 | 100.0 | 1695 | 21 | AAA69617  | Eucalyptus grandis |
| c 30 | 12 | 100.0 | 1722 | 21 | AAA57402  | cDNA sequence enco |
| c 31 | 12 | 100.0 | 1722 | 21 | AAA57439  | cDNA sequence enco |
| c 32 | 12 | 100.0 | 1947 | 22 | AAH76656  | Transducin 44 cDNA |
| c 33 | 12 | 100.0 | 2000 | 19 | AAV49651  | Human SRCR protea  |
| c 34 | 12 | 100.0 | 2261 | 23 | ABL27524  | Drosophila melanog |
| c 35 | 12 | 100.0 | 2305 | 22 | AAH15967  | Human cDNA sequenc |
| c 36 | 12 | 100.0 | 2671 | 21 | AAA57401  | Genomic sequence o |
| c 37 | 12 | 100.0 | 2671 | 21 | AAA57438  | Genomic sequence o |
| c 38 | 12 | 100.0 | 2674 | 23 | ABL09501  | Drosophila melanog |
| c 39 | 12 | 100.0 | 2678 | 21 | AAA57400  | Genomic sequence o |
| c 40 | 12 | 100.0 | 2678 | 21 | AAA57437  | Genomic sequence o |
| c 41 | 12 | 100.0 | 2993 | 22 | AAH18194  | Human cDNA sequenc |
| c 42 | 12 | 100.0 | 3107 | 23 | AAH69876  | DNA encoding novel |
| c 43 | 12 | 100.0 | 3116 | 16 | AAU02948  | EPH-like receptor  |
| c 44 | 12 | 100.0 | 3197 | 21 | AAC76409  | Human ORFX ORF1964 |
| c 45 | 12 | 100.0 | 3342 | 23 | AAH87118  | DNA encoding novel |

# ALIGNMENTS

RESULT 1  
ID AAT63580/c  
ID AAT63580 standard; DNA; 32 BP.  
AC AAT63580;  
XX  
XX  
DT 01-JUL-1997 (first entry)  
DE PCR primer used in CD59 minigene No. 1 construction.  
DE  
DE  
KW Xenotransplantation; organ transplant; transgenic animal;  
KW transgenic pig; transgenic mouse; antibody mediated rejection;  
KW hyperacute rejection; CD59; complement inhibitor; primer; PCR;  
KW polymerase chain reaction; minigene; ss.  
OS Synthetic.  
PN WO9712035-A2.  
XX  
XX  
PD 03-APR-1997.  
XX  
XX  
PF 23-SEP-1996; 96WO-US15255.  
XX  
PR 03-JUL-1996; 96US-0675773.  
PR 27-SEP-1995; 95US-0004461.  
XX  
PA (NEXT-) NEXTRAN.  
XX  
PI Byrne GW, Diamond LE, Logan JS, Sharma A;  
XX WPI; 1997-225881/20.  
XX

PT Transgenic animals expressing antigen reducing enzyme and complement  
 PT inhibitor - used for production of materials suitable for human  
 PT transplantation having a reduced risk of rejection

XX Example 1.1; Page 64; 146pp; English.

XX A sense primer (AAT63579) and antisense primer (AAT63580) were designed  
 CC to amplify a 1033 bp sequence of the 3' flanking region of the  
 CC human CD59 gene, with the introduction of a 5' SacI site and 3'  
 CC XhoI site. The PCR product was joined to CD59 cDNA contg. exons 2,  
 CC 3 and 4 plus 667 bp of 3' untranslated region (see also AAT63577-78)  
 CC to produce a CD59 minigene. This minigene was shown to retain  
 CC biological function when expressed in transgenic pigs and mice.  
 CC Expression of a complement inhibitor such as CD59 on endothelial  
 CC cells of transgenic animals can provide materials suitable for  
 CC transplantation to humans, suppressing complement activation and  
 CC reducing immune reaction.

XX Sequence 32 BP; 3 A; 9 C; 11 G; 9 T; 0 other;

Query Match 100.0%; Score 12; DB 18; Length 32;  
 Best Local Similarity 100.0%; Pred. No. 8.9e+02;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
 |||||

DB 17 CAGGAGCCTCGA 6

RESULT 2

ABAO3059/c

ID ABAO3059 standard; DNA; 58 BP.

AC ABAO3059;

XX 05-FEB-2002 (first entry)

XX Oligonucleotide #2 used to construct a HA fusion sequence.

XX Human; serum albumin; HA; antiinflammatory; immunosuppressive; cardiac;

XX neoptotic; neuroprotective; gene therapy; immune disorder; wound healing;

XX hyperproliferative disorder; renal disorder; cardiovascular disorder;

XX respiratory disorder; neurological disease; endocrine disorder;

XX reproductive system disorder; infectious disease;

XX gastrointestinal disorder; ss.

XX Homo sapiens.

XX WO200179444-A2.

XX 25-OCT-2001.

XX 12-APR-2001; 2001WO-US12013.

XX 12-APR-2000; 2000US-229358P.

XX 25-APR-2000; 2000US-199384P.

XX 21-DEC-2000; 2000US-256931P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Haseltine WA;

XX WPI; 2001-616755/71.

XX Albumin fusion proteins comprising a therapeutic protein and albumin,

PT useful in the treating immune system disorders (e.g. transplant

PT rejection), blood related disorders (e.g. myocardial infarction) and

PT hyperproliferative disorders -

XX Example 2; Page 465; 606pp; English.

XX The present invention relates to albumin fusion proteins, which comprise

CC a therapeutic protein and albumin. The albumin fusion proteins are useful  
 CC in the treatment, prevention, diagnosis, and/or detection of  
 CC diseases/disorders such as immune system disorders (e.g. transplant  
 CC rejection), blood related disorders (e.g. myocardial infarction),  
 CC hyperproliferative disorders (e.g. childhood acute myeloid leukemia),  
 CC renal disorders (e.g. glomerulonephritis), cardiovascular disorders (e.g.  
 CC arrhythmias), respiratory disorders (e.g. non-allergic rhinitis),  
 CC neurological diseases (e.g. Alzheimer's disease), endocrine disorders  
 CC (e.g. pheochromocytoma), reproductive system disorders (e.g. syphilis),  
 CC infectious diseases (e.g. measles), gastrointestinal disorders (e.g.  
 CC irritable bowel syndrome) and wound healing. In the present invention,  
 CC human serum albumin (HA; see AAM52567) was used to generate fusion  
 CC proteins. The present sequence was used to illustrate the invention.

XX Sequence 58 BP; 13 A; 15 C; 12 G; 18 T; 0 other;

Query Match 100.0%; Score 12; DB 22; Length 58;  
 Best Local Similarity 100.0%; Pred. No. 8.9e+02;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
 |||||

DB 51 CAGGAGCCTCGA 40

RESULT 3

AAD22289/c

ID AAD22289 standard; DNA; 59 BP.

XX AAD22289;

XX 12-FEB-2002 (first entry)

XX Human albumin (HA) mutagenic PCR primer #2.

XX Human; albumin; HA; fusion protein; immune system disorder; syphilis;

XX transplant rejection; blood related disorder; myocardial infarction;

XX hyperproliferative disorder; acute myeloid leukaemia; renal disorder;

XX glomerulonephritis; cardiovascular disease; arrhythmia; rhinitis;

XX respiratory disorder; neurological disease; Alzheimer's disease;

XX endocrine disorder; pheochromocytoma; reproductive system disorder;

XX measles; gastrointestinal disorder; irritable bowel syndrome; HIV;

XX human immunodeficiency virus; wound healing; renal cell carcinoma;

XX melanoma; gene therapy; PCR primer; ss.

XX Homo sapiens.

XX WO200179258-A1.

XX 25-OCT-2001.

XX 12-APR-2001; 2001WO-US12008.

XX 12-APR-2000; 2000US-229358P.

XX 25-APR-2000; 2000US-199384P.

XX 21-DEC-2000; 2000US-256931P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (PRIN-) PRINCIPALIA PHARM CORP.

XX Rosen CA, Sadeghi H, Prior CP, Turner AJ;

XX WPI; 2001-602931/68.

XX Albumin fusion proteins comprising a therapeutic protein and albumin,

PT useful in the treating metastatic renal cell carcinoma, metastatic

PT melanoma, malignant melanoma, renal cell carcinoma, HIV (human

PT immunodeficiency virus) or infection -

XX Example 2; Page 190; 325pp; English.

XX The invention relates to albumin fusion proteins comprising therapeutic



CC protein and human albumin (HA). The albumin fusion proteins are useful  
 CC in the treatment, prevention, diagnosis, and/or detection of diseases,  
 CC disorders such as immune system disorders (transplant rejection); blood  
 CC related disorders (myocardial infarction); hyperproliferative disorders  
 CC (childhood acute myeloid leukaemia); renal disorder (glomerulonephritis);  
 CC cardiovascular disorders (arrhythmias); respiratory disorders  
 CC (non-allergic rhinitis); neurological diseases (Alzheimer's disease);  
 CC endocrine disorders (pheochromocytoma); reproductive system disorders  
 CC (syphilis); infectious diseases (measles); gastrointestinal disorders  
 CC (irritable bowel syndrome) and wound healing. The albumin fusion  
 CC proteins are also used in the treatment of metastatic renal cell  
 CC carcinoma, metastatic melanoma, malignant melanoma and HIV (human  
 CC immunodeficiency virus) infection. Nucleic acid encoding albumin fusion  
 CC protein is useful in gene therapy. The present sequence is a mutagenic  
 CC PCR primer used for engineering the Xho and Cla I restriction sites  
 CC into the DNA encoding HA protein in pPC0006 plasmid.  
 CC Note: The present sequence shown in sequence listing of the specification  
 CC lacks a nucleotide base at its 3' end.  
 XX  
 SQ Sequence 59 BP; 13 A; 16 C; 12 G; 18 T; 0 other;

Query Match 100.0%; Score 12; DB 22; Length 59;

Best Local Similarity 100.0%; Pred. No. 8.9e+02; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12  
 |||||  
 Db 51 CAGGAGCCTCGA 40

#### RESULT 4

ID AAD21640/c  
 XX AAD21640;

DT 28-JAN-2002 (first entry)

XX Human albumin (HA) PCR primer #2.

XX Human; albumin; HA; fusion protein; therapeutic protein; vulnery;  
 KW immune system disorder; transplant rejection; blood related disorder;  
 KW myocardial infarction; hyperproliferative disorder; glomerulonephritis;  
 KW childhood acute myeloid leukemia; cardiovascular disorder; arrhythmia;  
 KW respiratory disorder; gene therapy; non-allergic rhinitis; nontropic;  
 KW neurological disease; Alzheimer's disease; reproductive system disorder;  
 KW endocrine disorder; pheochromocytoma; infectious disease; antiarthritic;  
 KW measles; gastrointestinal disorder; irritable bowel syndrome; syphilis;  
 KW wound healing; antiinflammatory; immunosuppressive; neuroprotective;  
 KW cardiant; cytostatic; antileukaemic; antirheumatic; antimicrobial;  
 KW renal disorder; PCR primer; ss.

XX Homo sapiens.

XX W0200179443-A2.

XX 25-OCT-2001.

XX 12-APR-2001; 2001WO-US11924.

XX 12-APR-2000; 2000US-229358P.

XX 25-APR-2000; 2000US-199384P.

XX 21-DEC-2000; 2000US-256931P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Haseltine WA;

XX WPI; 2001-616754/71.

XX Albumin fusion proteins comprising a therapeutic protein and albumin,

PT useful in the treating immune system disorders (e.g. transplant

PT rejection), blood related disorders (e.g. myocardial infarction) and  
 PT hyperproliferative disorders -  
 XX  
 PS Example 2; Page 231; 380pp; English.

XX The invention relates to albumin fusion proteins comprising therapeutic  
 CC protein and human albumin (HA). Therapeutic protein fused to albumin  
 CC have an extended shelf-life. The albumin fusion proteins are useful in  
 CC the treatment, prevention, diagnosis and/or detection of diseases,  
 CC disorders such as immune system disorders (e.g. transplant rejection),  
 CC blood related disorders (e.g. myocardial infarction), hyperproliferative  
 CC disorders (e.g. childhood acute myeloid leukaemia), renal disorders  
 CC (e.g. glomerulonephritis), cardiovascular disorders (e.g. arrhythmias),  
 CC respiratory disorders (e.g. non-allergic rhinitis), neurological  
 CC diseases (e.g. Alzheimer's disease), endocrine disorders (e.g.  
 CC pheochromocytoma), reproductive system disorders (e.g. syphilis),  
 CC infectious diseases (e.g. measles), gastrointestinal disorders (e.g.  
 CC irritable bowel syndrome) and wound healing. Nucleic acids encoding  
 CC albumin fusion protein is used in gene therapy. The present sequence  
 CC is a PCR primer used for engineering the Xho and Cla I restriction  
 CC sites into the DNA encoding HA protein in pPC0006 plasmid.  
 CC Note: This sequence SEQ.ID.NO.20 is stated to be similar to the  
 CC sequence shown in the sequence listing. However this sequence  
 CC contains an additional base 'C' at the 3' end which is absent in  
 CC the sequence shown in sequence listing.

XX Sequence 59 BP; 13 A; 16 C; 12 G; 18 T; 0 other;

Query Match 100.0%; Score 12; DB 22; Length 59;

Best Local Similarity 100.0%; Pred. No. 8.9e+02; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12  
 |||||  
 Db 51 CAGGAGCCTCGA 40

#### RESULT 5

AAD20007/c

ID AAD20007 standard; DNA; 59 BP.

XX AAD20007;

DT 18-DEC-2001 (first entry)

XX Human albumin (HA) PCR primer #2.

XX Human; albumin; HA; immune system disorder; transplant rejection;  
 KW blood related disorder; myocardial infarction; glomerulonephritis;  
 KW hyperproliferative disorder; childhood acute myeloid leukaemia;  
 KW renal cell carcinoma; cardiovascular disorder; antimicrobial; vulnery;  
 KW arrhythmia; respiratory disorder; non-allergic rhinitis; antileukaemic;  
 KW neurological disease; Alzheimer's disease; endocrine disorder; measles;  
 KW pheochromocytoma; reproductive system disorder; neuroprotective; syphilis;  
 KW infectious disease; gastrointestinal disorder; wound healing; nontropic;  
 KW irritable bowel syndrome; HIV; human immunodeficiency virus infection;  
 KW cytostatic; antiinflammatory; gene therapy; immunosuppressive; cardiant;  
 KW antiarthritic; antirheumatic; renal disorder; melanoma; PCR primer; ss.

XX Homo sapiens.

XX W0200179480-A1.

XX 25-OCT-2001.

XX 12-APR-2001; 2001WO-US11991.

XX 12-APR-2000; 2000US-229358P.

XX 25-APR-2000; 2000US-199384P.

XX 21-DEC-2000; 2000US-256931P.

XX (HUMA-) HUMAN GENOME SCI INC.

```

XX Rosen CA, Haseltine WA;
XX WPI; 2001-616756/71.
XX
XX Albumin fusion proteins comprising a therapeutic protein and albumin,
XX useful in the treating metastatic renal cell carcinoma, metastatic
XX melanoma, malignant melanoma, renal cell carcinoma, HIV (human
XX immunodeficiency virus) or infection -
XX
XX Example 2; Page 244; 394pp; English.
XX
XX The invention relates to human albumin (HA) fusion proteins and their
XX corresponding nucleic acid sequences. Therapeutic proteins fused to
XX albumin or its fragments have an extended shelf-life. The albumin
XX fusion proteins are useful in the treatment, prevention, diagnosis,
XX and/or detection of diseases, disorders such as immune system
XX disorders (e.g. transplant rejection), blood related disorders (e.g.
XX myocardial infarction), hyperproliferative disorders (e.g. childhood
XX acute myeloid leukaemia, metastatic renal cell carcinoma, metastatic
XX melanoma, malignant melanoma, renal cell carcinoma), renal disorders
XX (e.g. glomerulonephritis), cardiovascular disorders (e.g. arrhythmias),
XX respiratory disorders (e.g. non-allergic rhinitis), neurological
XX diseases (e.g. Alzheimer's disease), endocrine disorders (e.g.
XX pheochromocytoma), reproductive system disorders (e.g. syphilis),
XX infectious diseases (e.g. measles), gastrointestinal disorders (e.g.
XX irritable bowel syndrome), HIV (human immunodeficiency virus) infection
XX and wound healing. Nucleic acids encoding albumin fusion protein is
XX used in gene therapy. The present sequence is a mutagenic PCR primer
XX used for engineering the Xho and Cla I restriction sites into the
XX fusion leader sequence just 5' end of the DNA encoding the HA
XX protein in pPPC0006 plasmid.
XX Note: This sequence SEQ.ID.NO.20 is stated to be similar to the
XX sequence shown in the sequence listing. However this sequence
XX contains an additional base 'C' at the 3' end which is absent in
XX the sequence shown in sequence listing..
XX
XX Sequence 59 BP; 13 A; 16 C; 12 G; 18 T; 0 other;

Query Match 100.0%; Score 12; DB 22; Length 59;
Best Local Similarity 100.0%; Pred. No. 8.9e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctoga 12
Db 51 CAGGAGCCTCGA 40
|||||

RESULT 6
AAK51355
ID AAK51355 standard; DNA; 193 BP.
XX
XX AAK51355;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human bone marrow expressed single exon probe SEQ ID NO: 25912.
XX
XX Human: bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
XX Homo sapiens.
XX
XX WO200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00668.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 25-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.

Rosen CA, Haseltine WA;
WPI; 2001-616756/71.
(MOLE-) MOLECULAR DYNAMICS INC.
Penn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2001-488900/53.
Human genome-derived single exon nucleic acid probes useful for
analyzing gene expression in human bone marrow -
Example 4; SEQ ID NO: 25912; 658pp + Sequence Listing; English.
The present invention provides a number of single exon nucleic acid
probes which are derived from genomic sequences expressed in the human
bone marrow. They can be used to measure gene expression in bone marrow
samples, which may enable the improved diagnosis and treatment of cancers
such as lymphoma, leukaemia and myeloma. The present sequence is one of
the probes of the invention.
Sequence 193 BP; 32 A; 47 C; 27 G; 87 T; 0 other;

Query Match 100.0%; Score 12; DB 22; Length 193;
Best Local Similarity 100.0%; Pred. No. 8.8e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctoga 12
Db 63 caggagcctoga 74
|||||

RESULT 7
ABL27525
ID ABL27525 standard; DNA; 261 BP.
XX
XX ABL27525;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 34048.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 34048; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is

```

CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB157737-AB172072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 261 BP; 83 A; 66 C; 68 G; 44 T; 0 other;

Query Match 100.0%; Score 12; DB 23; Length 261;  
Best Local Similarity 100.0%; Pred. No. 8.8e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
Db 109 caggagcctcga 120  
|||||

RESULT 8  
AAC31784/C  
ID AAC31784 standard; cDNA; 292 BP.

XX AC AAC31784;

XX 06-OCR-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 35859.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
PS Claim 1: SEQ ID 35859; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.

XX Sequence 292 BP; 63 A; 76 C; 73 G; 80 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 292;  
Best Local Similarity 100.0%; Pred. No. 8.8e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
Db 222 CAGGAGCCTCGA 211  
|||||

RESULT 9  
AAA69618/C  
ID AAA69618 standard; cDNA; 315 BP.

XX AC AAA69618;

XX 08-NOV-2000 (first entry)

XX Eucalyptus grandis NADPH cDNA SEQ ID NO: 92.

XX Eucalyptus grandis; Pinus radiata; modification; isoprenoid; plant;  
KW metabolism; isoprenoid biosynthetic pathway; terpenoid; steroid;  
KW genome mapping; physical mapping; positional cloning; forestry;  
KW agriculture; medicine; fermentation; plant development; pest resistance;  
KW pinene; myrcene; Monterey pine; ss.

XX Eucalyptus grandis.

XX WO200036081-A2.

XX 22-JUN-2000.

XX 16-DEC-1999; 99WO-NZ00219.

XX 17-DEC-1998; 98US-0215504.

XX 29-JUL-1999; 99US-0146441.

XX (GENE-) GENESIS RES & DEV CORP LTD.

XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX Havukkala IJ;

XX WPI; 2000-431575/37.

XX P-PSDB; AAB18018.

XX New plant polynucleotides encoding polypeptides involved in the  
PT production and modification of isoprenoids, useful in forestry and  
PT agriculture for manipulation of isoprenoid metabolism -

XX Claim 1; Page 82; 164pp; English.

XX The present invention describes plant polynucleotides encoding  
CC polypeptides involved in the production and modification of isoprenoids,  
CC such as terpenoid and steroid compounds. The polynucleotides are used  
CC in genome mapping, in physical mapping and in positional cloning of  
CC genes. The polynucleotides and polypeptides are useful in forestry and  
CC agriculture for manipulation of isoprenoid metabolism, in medicine for  
CC therapeutic effects, including direct application in diseased organisms  
CC or indirect application by transgenic organisms and in fermentation and  
CC chemical processing industries involving isoprenoids. In plant  
CC applications, manipulating isoprenoid pathways or isoprenoid composition  
CC may, for example, affect plant development, pest resistance, and the  
CC value of extractives (e.g. pinene and myrcene). The ubiquitous and  
CC varied roles of isoprenoids make the polynucleotides attractive targets  
CC for biotechnical applications in a variety of fields. AAA69527 to  
CC AAA69690 and AAB18004 to AAB18143 represent Eucalyptus grandis and Pinus  
CC radiata polynucleotides and proteins used in the exemplification of the  
CC present invention.

XX Sequence 315 BP; 77 A; 96 C; 90 G; 52 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 315;  
 Best Local Similarity 100.0%; Pred. No. 8.8e+02;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
 |||||  
 DB 120 CAGGAGCCTCGA 109

RESULT 10  
 AAC75741/c  
 ID AAC75741 standard; cDNA; 341 BP.  
 XX  
 AC AAC75741;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORFX ORF1296 polynucleotide sequence SEQ ID NO:2591.  
 XX  
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200058473-A2.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 31-MAR-2000; 2000WO-US08621.  
 XX  
 PR 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shinkets RA, Leach M;  
 XX  
 DR WPI; 2000-602362/57.  
 DR  
 DR P-PSDB; AAB41532.  
 XX  
 XX Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 XX Claim 5; Page 1852; 5507pp; English.  
 XX  
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianaemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,

CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus, lupus  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 SQ Sequence 341 BP; 74 A; 82 C; 125 G; 60 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 341;  
 Best Local Similarity 100.0%; Pred. No. 8.8e+02;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
 |||||  
 DB 267 CAGGAGCCTCGA 256

RESULT 11  
 AAC04396  
 ID AAC04396 standard; cDNA; 394 BP.  
 XX  
 AC AAC04396;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein 5' EST, SEQ ID NO: 8471.  
 XX  
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1033401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000; 2000EP-0200610.  
 XX  
 PR 26-FEB-1999; 99US-0122487.  
 XX  
 PA (GEST) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 DR WPI; 2000-500381/45.  
 XX  
 DR New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 XX Claim 1; SEQ ID 8471; 71pp + CD-ROM; English.  
 PS  
 XX The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.  
 XX  
 SQ Sequence 394 BP; 86 A; 110 C; 122 G; 73 T; 3 other;

Query Match 100.0%; Score 12; DB 21; Length 394;  
Best Local Similarity 100.0%; Pred. No. 8.8e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
|||||  
DB 117 caggagcctcga 128

RESULT 12  
AAC08874/c  
ID AAC08874 standard; cDNA; 400 BP.  
XX AC AAC08874;  
XX 06-OCT-2000 (first entry)  
DE Human secreted protein 5' EST, SEQ ID NO: 12949.  
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.  
XX Homo sapiens.  
XX EP1033401-A2.  
XX 06-SEP-2000.  
XX 21-FEB-2000; 2000EP-0200610.  
XX 26-FEB-1999; 99US-0122487.  
XX (GBST ) GENSET.  
XX Dunas Milne Edwards J, Duclert A, Giordano J;  
XX WPI; 2000-500381/45.  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX Claim 1; SEQ ID 12949; 71pp + CD-ROM; English.  
XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.  
XX Sequence 400 BP; 113 A; 137 C; 106 G; 43 T; 1 other;

Query Match 100.0%; Score 12; DB 21; Length 400;  
Best Local Similarity 100.0%; Pred. No. 8.8e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
|||||  
DB 49 CAGGAGCCTCGA 38

RESULT 13

AAT74033  
ID AAT74033 standard; cDNA; 450 BP.  
XX AC AAT74033;  
XX 15-SEP-1997 (first entry)  
DE Aloe arborescens phosphoenolpyruvate carboxylase probe.  
XX PEPCase; monocotyledonous CAM plant; carbon dioxide fixation;  
KW transformation; polymerase chain reaction; ss.  
XX Aloe arborescens.  
XX OS  
XX PN JP09107975-A.  
XX 28-APR-1997.  
XX 19-OCT-1995; 95JP-0294986.  
XX 19-OCT-1995; 95JP-0294986.  
XX (MITK ) MITSUI TOATSU CHEM INC.  
XX WPI; 1997-292468/27.  
XX Phospho:enol:pyruvate carboxylase gene of a monocotyledonous CAM  
PT plant - used to transform plants so that they can fix carbon  
PT di:oxide in dry conditions  
XX Example 1; Page 10; 12pp; Japanese.  
XX The present sequence represents a probe for phosphoenolpyruvate  
CC carboxylase (PEPCase) from Aloe arborescens. A plasmid containing  
CC the DNA sequence of PEPCase can be introduced into a monocotyledonous  
CC plant to enhance PEPCase activity. The transformed plant has a high  
CC ability to fix carbon dioxide under dry conditions. It can also  
CC increase corn productivity when introduced into a corn plant.  
XX Sequence 450 BP; 96 A; 101 C; 147 G; 106 T; 0 other;

Query Match 100.0%; Score 12; DB 18; Length 450;  
Best Local Similarity 100.0%; Pred. No. 8.8e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
|||||  
DB 424 caggagcctcga 435

RESULT 14  
AAZ40844/c  
ID AAZ40844 standard; DNA; 475 BP.  
XX AC AAZ40844;  
XX 18-JAN-2000 (first entry)  
DE Secreted protein EST coding sequence 78-6-2-B10-FL1.  
XX Secreted protein; fingerprint identification technique;  
KW chromosome mapping; human; hereditary disease; diagnosis; cancer;  
KW hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;  
KW autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;  
KW renal injury; amino aciduria; hypoglycaemia; male rat infertility;  
KW hypertension; ss.  
XX Homo sapiens.  
XX WO9940189-A2.  
XX 12-AUG-1999.

XX 09-FEB-1999; 99WO-IB00282.  
 XX 09-FEB-1998; 98US-0074121.  
 PR 13-APR-1998; 98US-0081563.  
 PR 10-AUG-1998; 98US-0096116.  
 PR 04-SEP-1998; 98US-0099273.  
 XX (GEST ) GENSET.  
 XX Bouqueloret L, Duclert A, Dumas Milne Edwards J;  
 PI WPI; 1999-600966/51.  
 XX P-PSDB; AAY59716.  
 XX Extended cDNAs useful for expressing secreted proteins and to obtain  
 PT specific antibodies -  
 XX Claim 1; Page 224-225; 244pp; English.  
 XX This sequence encodes a human secreted protein of the invention. The  
 CC extended cDNAs (or genomic DNAs obtainable from them) may be used to  
 CC prepare PCR primers and probes. These are useful for forensic matching or  
 CC positive identification by DNA sequencing. They may also be used in  
 CC alternative fingerprint identification techniques. Antibodies against the  
 CC proteins encoded by the extended cDNAs are useful in identification of  
 CC tissue types or cell species, as well as identifying tissue specific  
 CC soluble proteins. The sequences can be used for chromosome mapping and  
 CC identification of genes associated with hereditary diseases or drug  
 CC response. Signal sequences from the cDNAs can be used in construction of  
 CC secretion vectors. Other sequences derived from the extended cDNAs can be  
 CC used to clone upstream genomic DNA sequences including promoters. This is  
 CC in turn useful for identifying proteins that interact with promoter  
 CC sequences. Some of the proteins may be useful in diagnosing and treating  
 CC several disorders including, but not limited to: cancer, hyperlipidaemia,  
 CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and  
 CC rheumatic diseases, embryogenic disorders, hypertension, renal injury,  
 CC amino acidurias, hypoglycaemia, male rat infertility and myopathies.  
 XX Sequence 475 BP; 121 A; 118 C; 112 G; 121 T; 3 other;  
 SQ

Query Match 100.0%; Score 12; DB 20; Length 475;  
 Best Local Similarity 100.0%; Pred. No. 8.8e+02;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
 |||||  
 Db 223 CAGGAGCCTCGA 212

RESULT 15  
 AAH97774/C  
 ID AAH97774 standard; DNA; 552 BP.  
 XX AC AAH97774;  
 XX 10-OCT-2001 (first entry)  
 XX Murine 7-transmembrane G-protein coupled receptor coding sequence #18.  
 XX Murine; hematopoietic stem cell; signalling; vaccine; 7TM-GPCR;  
 XX 7-transmembrane G-protein coupled protein receptor; ds.  
 XX Mus sp.  
 XX WO200160999-A1.  
 XX 23-AUG-2001.  
 XX 14-FEB-2001; 2001WO-US04700.  
 XX 14-FEB-2000; 2000US-0182377.

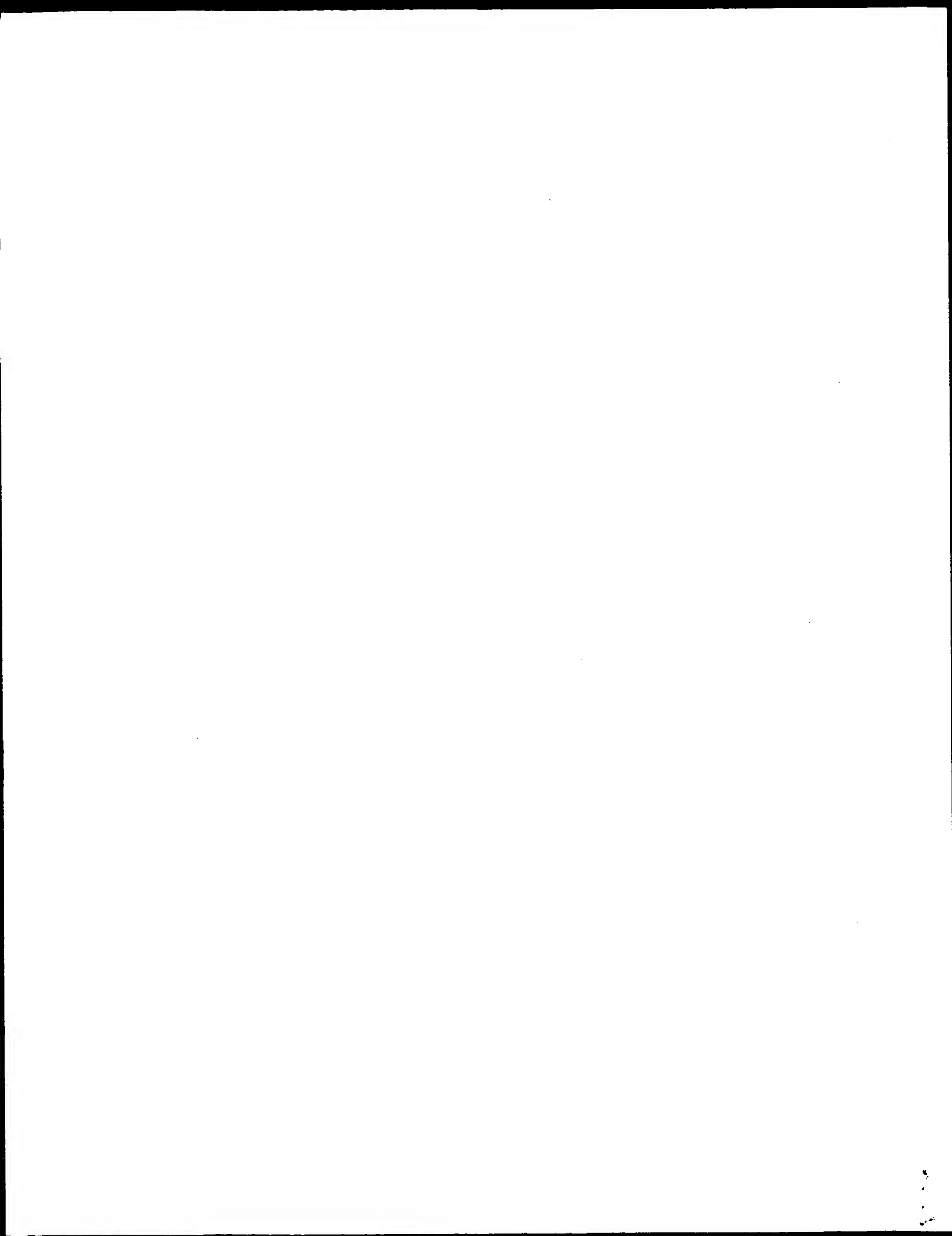
XX (JMCL-) IMCLONE SYSTEMS INC.  
 PA (UYPR-) UNIV PRINCETON.  
 XX Lemischka IR, Witte L, Pereira DS;  
 XX WPI; 2001-522596/57.  
 XX DNA Sequences encoding 7-transmembrane G-protein coupled protein  
 PT receptors characteristic of hematopoietic stem cells, useful for  
 PT treating leukemia -  
 XX Claim 1; Page 30-31; 176pp; English.  
 XX The present invention relates to murine coding sequences for  
 CC 7-transmembrane G-protein coupled protein receptors (7TM-GPCRs). The  
 CC present sequence is one such murine 7TM-GPCR coding sequence. The present  
 CC sequence was derived from hematopoietic stem cells. The present sequence  
 CC and its corresponding protein are useful in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate 7TM-GPCR expression.  
 CC 7TM-GPCRs identify specific signalling molecules, to activate an  
 CC effector-signalling cascade that triggers an intracellular response and  
 CC eventually a biological effect.  
 XX Sequence 552 BP; 126 A; 120 C; 125 G; 95 T; 86 other;  
 SQ

Query Match 100.0%; Score 12; DB 22; Length 552;  
 Best Local Similarity 100.0%; Pred. No. 8.8e+02;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
 |||||  
 Db 130 CAGGAGCCTCGA 119

Search completed: July 31, 2002, 14:11:24  
 Job time: 17381 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:01:53 ; Search time 6034.22 Seconds  
(without alignments)  
41.616 Million cell updates/sec

Title: US-09-899-718A-6  
Perfect score: 12  
Sequence: 1 caggagcctcga 12

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_hg: \*  
3: gb\_in: \*  
4: gb\_cm: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pi: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*  
29: em\_vi: \*  
30: em\_htg\_hum: \*  
31: em\_htg\_inv: \*  
32: em\_htg\_other: \*  
33: em\_htgo\_inv: \*

ALIGNMENTS

RESULT 1

AX349068

LOCUS

DEFINITION

AX349068

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

AX349068 12 bp DNA linear PAT 06-FEB-2002  
Sequence 6 from Patent WO0202785.

AX349068.1 GI:18615103

Synthetic construct.

Synthetic construct

artificial sequence.

1 (sites)

Sprunck, S., Kluth, A., Becker, D., Luetticke, S. and Loerz, H.

Promoters of gene expression in plant caryopses

Patent: WO 0202785-A 6 10-JAN-2002;

Avantis CropScience GmbH (DE)

Location/Qualifiers

1..12

/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="Oligonucleotide"

3 a 4 c 4 g 1 t

BASE COUNT

ORIGIN

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
|------------|-------|-------------|--------|----|-------------|

```

Query Match          100.0%; Score 12; DB 6; Length 12;
Best Local Similarity 100.0%; Pred. No. 7.9e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12
|||||
Db 1 CAGGAGCCTCGA 12

RESULT 2
AF150645
LOCUS      Petrotux sanguineus 218 bp DNA linear VRT 22-AUG-2000
DEFINITION Petrotux sanguineus 28S ribosomal RNA gene, partial sequence.
ACCESSION  AF150645
VERSION     AF150645.1 GI:4960098
KEYWORDS   .
SOURCE     Petrotux sanguineus.
ORGANISM   Petrotux sanguineus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Paracanthopterygii; Ophidiiformes; Ophidiidae;
            Petrotux.
REFERENCE   1 (bases 1 to 218)
AUTHORS    Wiley,E.O., David Johnson,G. and Wheaton Dimmick,W.
TITLE      The interrelationships of Acanthomorph fishes: A total evidence
            approach using molecular and morphological data
JOURNAL    Biochemical systematics and ecology. 28 (4), 319-350 (2000)
PUBMED     10725591
REFERENCE   2 (bases 1 to 218)
AUTHORS    Wiley,E.O., Johnson,G.D. and Dimmick,W.W.
TITLE      Direct Submission
JOURNAL    Submitted (12-MAY-1999) Natural History Museum, University of
            Kansas, Lawrence, KS 66045, USA
FEATURES   Location/Qualifiers
            source          1..218
                        /organism="Petrotux sanguineus"
                        /specimen_voucher="KU 22948"
                        /db_xref="taxon:94932"
                        <1..>218
                        /product="28S ribosomal RNA"
BASE COUNT 45 a 57 c 86 g 30 t

rRNA
BASE COUNT 45 a 57 c 86 g 30 t
ORIGIN
|||||

Query Match          100.0%; Score 12; DB 5; Length 218;
Best Local Similarity 100.0%; Pred. No. 5.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12
|||||
Db 43 CAGGAGCCTCGA 54

RESULT 3
AF150652
LOCUS      Mastacembelus sp. 218 bp DNA linear VRT 22-AUG-2000
DEFINITION Mastacembelus sp. 28S ribosomal RNA gene, partial sequence.
ACCESSION  AF150652
VERSION     AF150652.1 GI:4960105
KEYWORDS   .
SOURCE     Mastacembelus sp.
ORGANISM   Mastacembelus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percormorpha; Synbranchiformes;
            Mastacembelidae; Mastacembelus.
REFERENCE   1 (bases 1 to 218)
AUTHORS    Wiley,E.O., David Johnson,G. and Wheaton Dimmick,W.
TITLE      The interrelationships of Acanthomorph fishes: A total evidence
            approach using molecular and morphological data
JOURNAL    Biochemical systematics and ecology. 28 (4), 319-350 (2000)
PUBMED     10725591

```

```

REFERENCE           2 (bases 1 to 218)
AUTHORS            Wiley,E.O., Johnson,G.D. and Dimmick,W.W.
TITLE              Direct Submission
JOURNAL            Submitted (12-MAY-1999) Natural History Museum, University of
            Kansas, Lawrence, KS 66045, USA
FEATURES           Location/Qualifiers
            source          1..218
                        /organism="Mastacembelus sp."
                        /specimen_voucher="KU 22982"
                        /db_xref="taxon:94235"
                        <1..>218
                        /product="28S ribosomal RNA"
BASE COUNT        45 a 56 c 85 g 32 t
ORIGIN
|||||

rRNA
BASE COUNT        45 a 56 c 85 g 32 t
ORIGIN
|||||

Query Match        100.0%; Score 12; DB 5; Length 218;
Best Local Similarity 100.0%; Pred. No. 5.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12
|||||
Db 43 CAGGAGCCTCGA 54

RESULT 4
TSITSAA1/c
LOCUS      TSTTSAA1          258 bp DNA linear PLN 29-JUN-1999
DEFINITION Tamarix sp. Schultheis 19-94 internal transcribed spacer 1,
            complete sequence.
ACCESSION  AF084320
VERSION     AF084320.1 GI:5257020
KEYWORDS   .
SEGMENT    1 of 2
SOURCE     Tamarix sp. Schultheis 19-94.
ORGANISM   Tamarix sp. Schultheis 19-94
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Caryophyllidae; Caryophyllales; Tamaricaceae; Tamarix.
REFERENCE   1 (bases 1 to 258)
AUTHORS    Schultheis,L.M. and Baldwin,B.G.
TITLE      Molecular phylogenetics of Fouquieriaceae: evidence from nuclear
            rDNA ITS studies
JOURNAL    Am. J. Bot. 86, 578-589 (1999)
REFERENCE   2 (bases 1 to 258)
AUTHORS    Schultheis,L.M. and Baldwin,B.G.
TITLE      Direct Submission
JOURNAL    Submitted (14-AUG-1998) Integrative Biology, University of
            California, Berkeley, 3060 VLSB, Berkeley, CA 94720, USA
FEATURES           Location/Qualifiers
            source          1..258
                        /organism="Tamarix sp. Schultheis 19-94"
                        /db_xref="taxon:98320"
                        /note="ITS1"
            misc_RNA        1..258
                        /product="internal transcribed spacer 1"
BASE COUNT        57 a 83 c 74 g 44 t
ORIGIN
|||||

Query Match        100.0%; Score 12; DB 8; Length 258;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12
|||||
Db 99 CAGGAGCCTCGA 88

RESULT 5
G49233
LOCUS      G49233          442 bp mRNA linear STS 31-AUG-1999
DEFINITION D1ErtD754e Mouse eight-cell stage embryo cDNA library (M. Ko) Mus

```

musculus STS cDNA clone J0521C08 3', sequence tagged site.  
 G49233  
 VERSION  
 G49233.1 GI:4726066  
 STS  
 house mouse.  
 SOURCE  
 Mus musculus  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 REFERENCE  
 1 (bases 1 to 442)  
 AUTHORS  
 Bergstrom,D.E., Gagnon,L.H. and Eicher,E.M.  
 TITLE  
 Genetic and physical mapping of the dreher locus on mouse  
 chromosome 1  
 JOURNAL  
 Genomics 59 (3), 291-299 (1999)  
 MEDLINE  
 99375322  
 PUBMED  
 10444330  
 COMMENT

Contact: Eva M. Eicher  
 The Eicher Laboratory  
 The Jackson Laboratory  
 600 Main Street, Bar Harbor, ME 04609 USA  
 Tel: 207/288-6344  
 Fax: 207/288-6077  
 Email: emee@jax.org  
 Primer A: GCACCTACGGTAATCAAGGC  
 Primer B: ACCTGAGTGATGTTGGCT  
 STS size: 200  
 PCR Profile:

Presoak: 97.0 degrees C for 30 seconds  
 Denaturation: 94.0 degrees C for 15 seconds  
 Annealing: 55.0 degrees C for 30 seconds  
 Polymerization: 72.0 degrees C for 30 seconds + 1 sec/cyc  
 Postsoak: 72.0 degrees C for 10 minutes  
 PCR Cycles: 40

Protocol:  
 Template: 25 ng  
 Primer: each 0.25 uM  
 dNTPs: each 200 uM  
 Taq polymerase: 0.025 units/ul  
 Total volume: 20 ul

Buffer:  
 MgCl2: 1.5 mM  
 KCl: 50 mM  
 Tris-HCl: 10 mM  
 pH: 8.3

STS designed from Mouse Genome Database DNA segment DI8td754e.

FEATURES  
 source  
 Location/Qualifiers  
 1..442  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="J0521C08"  
 /clone\_lib="Mouse eight-cell stage embryo cDNA library (M.  
 ko)"  
 /dev\_stage="eight-cell stage embryo"  
 23..222  
 primer\_bind  
 23..42  
 primer\_bind  
 complement(203..222)  
 BASE COUNT 136 a 90 c 140 g 75 t 1 others  
 ORIGIN

Query Match 100.0%; Score 12; DB 11; Length 442;  
 Best Local Similarity 100.0%; Pred. No. 4.8e+03;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12  
 |||||  
 Db 374 CAGGAGCCTCGA 385

RESULT 6

E12962  
 LOCUS  
 DEFINITION  
 CDNA encoding fragment of phosphoenolpyruvate carboxylase from  
 Aloe.  
 E12962  
 VERSION  
 E12962.1 GI:3251791  
 JP 1997107975-A/4.  
 KEYWORDS  
 unidentified.  
 ORGANISM  
 unidentified.  
 unclassified.  
 1 (bases 1 to 450)  
 REFERENCE  
 1 (bases 1 to 450)  
 AUTHORS  
 Shimada,H., Okamoto,T., Honda,H. and Fujimura,T.  
 TITLE  
 PEP-ASE GENE OF MONOCOTYLEDON CAM PLANT  
 JOURNAL  
 Patent: JP 1997107975-A 4 28-APR-1997;  
 MITSUI TOATSU CHEM INC  
 COMMENT  
 OS Aloe arborescens  
 PN JP 1997107975-A/4  
 PD 28-APR-1997  
 PF 19-OCT-1995 JP 1995294986  
 PI SHIMADA HIROAKI, OKAMOTO TOMOKO, HONDA HIDEO, FUJIMURA TATSUTO  
 PC C12N15/09,A01H5/00,C07H21/04,C12N1/21,C12N5/10,C12N9/86, PC  
 (C12N1/21,  
 PC C12R1:19),(C12N5/10,C12R1:91);  
 CC strandedness: Double;  
 CC topology: Linear;  
 CC hypothetical: No;  
 CC anti-sense: No;  
 FH Key  
 FH Location/Qualifiers  
 FT source  
 1..450  
 /organism="Aloe arborescens"  
 /tissue\_type="green leaf"  
 FT mat\_peptide  
 1..450  
 /product="Phosphoenolpyruvate carboxylase from  
 Aloe"  
 FT  
 FT Location/Qualifiers  
 1..450  
 /organism="unidentified"  
 /db\_xref="taxon:32644"

BASE COUNT 96 a 101 c 147 g 106 t  
 ORIGIN

Query Match 100.0%; Score 12; DB 6; Length 450;  
 Best Local Similarity 100.0%; Pred. No. 4.8e+03;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12  
 |||||  
 Db 424 CAGGAGCCTCGA 435

RESULT 7  
 HS301F5  
 LOCUS  
 DEFINITION  
 H.sapiens chromosome 22 CpG island DNA genomic MseI fragment, clone  
 301f5, complete read.  
 ACCESSION  
 Z79832  
 VERSION  
 Z79832.1 GI:1531822  
 KEYWORDS  
 Chromosome 22; CpG island.  
 SOURCE  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE  
 1 (bases 1 to 453)  
 AUTHORS  
 Clark,V.H., Cross,S.H., Simmen,M.W., Langford,C., Carter,N.,  
 Bickmore,W. and Bird,A.P.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (02-SEP-1996) ICMB, University of Edinburgh, King's  
 Buildings, Mayfield Rd, Edinburgh EH9 3UR, UK. E-mail contact:  
 VCLARK@srv0.bio.ed.ac.uk  
 2 (bases 1 to 453)

Query Match 100.0%; Score 12; DB 6; Length 450;  
 Best Local Similarity 100.0%; Pred. No. 4.8e+03;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 6

```

AUTHORS      Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.
TITLE        Purification of CpG islands using a methylated DNA binding column
JOURNAL      Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE      94282070
COMMENT      Vector: pGEM-5zf(-).
FEATURES     Location/Qualifiers
             1..453
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="sorted chromosome 22"
             /cell_line="lymphoblastoid"
             /clone_lib="CGI22.1"
             /clone="301f5"
BASE COUNT   99 a 120 c 148 g 81 t 5 others
ORIGIN
Query Match  100.0%; Score 12; DB 9; Length 453;
Best Local Similarity 100.0%; Pred. No. 4.8e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
    |||||
Db 133 CAGGAGCCTCGA 144

RESULT 8
HS302F12     453 bp DNA linear PRI 12-SEP-1996
LOCUS        H.sapiens chromosome 22 CpG island DNA genomic MseI fragment, clone
302f12, complete read.
ACCESSION    Z79864
VERSION      Z79864.1 GI:1531858
KEYWORDS     Chromosome 22; CpG island.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 453)
AUTHORS      Clark, V.H., Cross, S.H., Simmen, M.W., Langford, C., Carter, N.,
             Bickmore, W. and Bird, A.P.
TITLE        Direct Submission
JOURNAL      Submitted (02-SEP-1996) ICMB, University of Edinburgh, King's
             Buildings, Mayfield Rd, Edinburgh EH9 3JR, UK. E-mail contact:
             VCLARK@rsrv0.bio.ed.ac.uk
REFERENCE    2 (bases 1 to 453)
AUTHORS      Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.
JOURNAL      Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE      94282070
COMMENT      Vector: pGEM-5zf(-).
FEATURES     Location/Qualifiers
             1..453
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="sorted chromosome 22"
             /cell_line="lymphoblastoid"
             /clone_lib="CGI22.1"
             /clone="302f12"
BASE COUNT   104 a 122 c 147 g 79 t 1 others
ORIGIN

Query Match  100.0%; Score 12; DB 9; Length 453;
Best Local Similarity 100.0%; Pred. No. 4.8e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
    |||||
Db 132 CAGGAGCCTCGA 143

RESULT 9

```

```

ARI77408/c
LOCUS        ARI77408
DEFINITION   Sequence 147 from patent US 6312922.
ACCESSION    ARI77408
VERSION      ARI77408.1 GI:17919763
KEYWORDS     .
SOURCE       Unknown.
ORGANISM     Unclassified.
             1 (bases 1 to 475)
REFERENCE    Edwards, J.-B., Dumas, Milne., Duclert, A. and Bougueleret, L.
AUTHORS      Complementary DNAs
TITLE        Patent: US 6312922-A 147 06-NOV-2001;
JOURNAL      Location/Qualifiers
FEATURES     1..475
             /organism="unknown"
BASE COUNT   121 a 118 c 112 g 121 t 3 others
ORIGIN

Query Match  100.0%; Score 12; DB 6; Length 475;
Best Local Similarity 100.0%; Pred. No. 4.7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
    |||||
Db 223 CAGGAGCCTCGA 212

RESULT 10
AB02081S28/c
LOCUS        AB02081S28
DEFINITION   Homo sapiens gene for DMBT1, exon 28.
ACCESSION    AB020839
VERSION      AB020839.1 GI:4996264
KEYWORDS     DMBT1.
SEGMENT      28 of 40
SOURCE       Homo sapiens DNA.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (sites)
AUTHORS      Takeshita, H., Sato, M., Shiwaku, H.O., Samba, S., Sakurada, A.,
             Hoshi, M., Hayashi, Y., Tagawa, Y., Ayabe, H. and Horii, A.
TITLE        Expression of the DMBT1 gene is frequently suppressed in human lung
             cancer
JOURNAL      Jpn. J. Cancer Res. 90 (9), 903-908 (1999)
MEDLINE      20017478
REFERENCE    2 (bases 1 to 498)
AUTHORS      Horii, A.
TITLE        Direct Submission
JOURNAL      Submitted (06-DEC-1998) Akira Horii, Tohoku University School of
             Medicine, Department of Molecular Pathology; 2-1 Seiryomachi,
             Aoba-ku, Sendai, Miyagi 980-8575, Japan
             (E-mail: horii@mail.cc.tohoku.ac.jp, Tel: 81-22-717-8042,
             Fax: 81-22-717-8047)
FEATURES     Location/Qualifiers
             1..498
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /gene="DMBT1"
             /number=27
             92..415
             /gene="DMBT1"
             /number=28
             /product="DMBT1"
             416..>498
             /gene="DMBT1"
             /number=28
BASE COUNT   98 a 125 c 154 g 121 t
ORIGIN

```

Query Match 100.0%; Score 12; DB 9; Length 498;  
 Best Local Similarity 100.0%; Pred. No. 4.7e+03; Indels 0; Gaps 0;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
 |||||  
 Db 176 CAGGAGCCTCGA 165

RESULT 11  
 G39052  
 LOCUS 598 bp DNA linear STS 01-FEB-2001  
 DEFINITION Z11764 zebrafish AB Danio rerio STS genomic, sequence tagged site.  
 ACCESSION G39052  
 VERSION G39052.1 GI:3358261  
 KEYWORDS STS.  
 SOURCE zebrafish.  
 ORGANISM Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.  
 REFERENCE 1 (bases 1 to 598)  
 AUTHORS Shimoda,N., Knapik,E.W., Ziniti,J., Sim,C., Yamada,E., Kaplan,S.,  
 Jackson,D., de Sauvage,F., Jacob,H. and Fishman,M.C.  
 TITLE Zebrafish genetic map with 2000 microsatellite markers  
 JOURNAL Genomics 58 (3), 219-232 (1999)  
 MEDLINE 93033552  
 PUBMED 10373319

COMMENT  
 Contact: Mark C. Fishman  
 Cardiovascular Research Center  
 Massachusetts General Hospital  
 Mail code 1494100A, 149 13th Street, Charlestown, MA 02129, USA  
 Fax: 6177263806  
 Email: fishman@mgh.cvr.harvard.edu  
 http://zebrafish.mgh.harvard.edu  
 Primer A: CGACGCTTTGTGATGTTT  
 Primer B: CATCTGATTGCTTGTCTCCA  
 STS size: 297  
 PCR Profile:  
 Presoak: 94 degrees C for 5.0 minutes  
 Denaturation: 94 degrees C for 1.0 minute  
 Annealing: 58 degrees C for 1.0 minute  
 Polymerization: 72 degrees C for 1.5 minute  
 PCR Cycles: 27  
 Thermal Cycler: MJ Research PTC-100

Protocol:  
 Template: 10 ng  
 Primer: each 375 nM  
 dNTPs: each 200 uM  
 Taq Polymerase: 0.034 units/ul  
 Total Vol: 10 ul

Buffer:  
 MgCL2: 1.5 mM  
 KCL: 50 mM  
 Tris-HCL: 10 mM  
 PH: 8.3

Primers are available from Research Genetics Inc.  
 (http://www.resgen.com phone: 800-533-4363).

FEATURES  
 source

1. 598  
 /organism="Danio rerio"  
 /strain="AB"  
 /db\_xref="taxon:7955"  
 /sex="F"  
 /clone\_lib="zebrafish AB"  
 /dev\_stage="Adult"  
 /lab\_host="DH5alpha/IQ"  
 /note="vector: ml3mp19 with added BstXI site; V-type:  
 Phage; Genomic DNA from a single adult zebrafish of AB

strain was digested with AluI, Cac8I, HaeIII, NlaVI, or  
 RsaI. Fragments in the range of 250-500 bp were gel  
 purified and a BstXI linker was added. The fragments were  
 cloned into a modified M13mp19 vector and transformed  
 into E. Coli DH5alpha. Microsatellite sequences were  
 screened with labeled d(CA)15 and d(GT)15 oligonucleotide  
 probes."

STS  
 primer\_bind 123..419  
 primer\_bind complement(400..419)  
 BASE COUNT 137 a 149 c 142 g 155 t 15 others  
 ORIGIN

Query Match 100.0%; Score 12; DB 11; Length 598;  
 Best Local Similarity 100.0%; Pred. No. 4.6e+03;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
 |||||  
 Db 82 CAGGAGCCTCGA 93

RESULT 12  
 AF321195  
 LOCUS 680 bp DNA linear BCT 24-JAN-2001  
 DEFINITION Uncultured bacterium rcr\_15 16S ribosomal RNA gene, partial  
 sequence.  
 ACCESSION AF321195  
 VERSION AF321195.1 GI:12407991  
 KEYWORDS uncultured bacterium rcr\_15.  
 SOURCE uncultured bacterium rcr\_15  
 ORGANISM Bacteria; environmental samples.  
 REFERENCE 1 (bases 1 to 680)  
 AUTHORS Boomer,S.M., Dutton,B.E. and Lodge,D.P.  
 TITLE Diversity of novel green non-sulfur bacteria from Yellowstone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 680)  
 AUTHORS Boomer,S.M., Dutton,B.E. and Lodge,D.P.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-NOV-2000) Dept. of Biology, Western Oregon  
 University, 345 Monmouth Ave., Monmouth, OR 97361, USA  
 FEATURES  
 Location/Qualifiers  
 1..680  
 /organism="uncultured bacterium rcr\_15"  
 /db\_xref="taxon:148410"  
 /clone="rcr\_15"  
 /country="USA: Rabbit Creek, Yellowstone"  
 <1..>680  
 /product="16S ribosomal RNA"

BASE COUNT 194 a 176 c 192 g 118 t  
 ORIGIN

Query Match 100.0%; Score 12; DB 1; Length 680;  
 Best Local Similarity 100.0%; Pred. No. 4.5e+03;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12  
 |||||  
 Db 502 CAGGAGCCTCGA 513

RESULT 13  
 HUMSYBIAL/c  
 LOCUS Human synaptobrevin 1 (SYBI) gene, exon 1.  
 DEFINITION M36196 J05611  
 ACCESSION M36196.1 GI:338619  
 VERSION 1  
 KEYWORDS intrinsic membrane protein; synaptic vesicle protein 2A;  
 synaptobrevin 1.  
 SEGMENT 1 of 5

SOURCE Human DNA, clone pcVP18-[1,5].

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 727)

TITLE Archer, B.T. III, Ozcelik, T., Jahn, R., Francke, U. and Sudhof, T.C. Structures and chromosomal localizations of two human genes encoding synaptobrevins 1 and 2

JOURNAL J. Biol. Chem. 265 (28), 17267-17273 (1990)

MEDLINE 91009161

COMMENT Draft entry and computer-readable sequence for [J. Biol. Chem. (1990) in press] kindly submitted by T.C.Sudhof, 06-JUL-1990.

FEATURES

source 1..727

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/map="12p"

TATA\_signal 544..549

exon <721..722

/gene="SVB1"

/note="synaptobrevin 1 (SVB1); G00-125-292"

/number=1

intron 723..727

/gene="SVB1"

/note="synaptobrevin 1 intron A"

BASE COUNT 130 a 217 c 257 g 123 t

ORIGIN Chromosome 12.

Query Match 100.0%; Score 12; DB 9; Length 727;

Best Local Similarity 100.0%; Pred. No. 4.4e+03;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12

|||||

Db 84 CAGGAGCCTCGA 73

RESULT 14

G41958/c

LOCUS G41958 761 bp DNA linear STS 30-SEP-1998

DEFINITION SHGC-68982 Human Homo sapiens STS genomic, sequence tagged site.

ACCESSION G41958

VERSION G41958.1 GI:3668291

KEYWORDS STS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 761)

TITLE Myers, R.M.

JOURNAL Human STSs (1998)

COMMENT Unpublished

Contact: Richard M. Myers

Stanford Human Genome Center (SHGC)

Stanford University School of Medicine

Department of Genetics, M-344, Stanford, CA 94305, USA

Tel: 4157259687

Fax: 4157259689

Email: myers@shgc.stanford.edu

Primer A: CAGGCACAAAGACTACCCACTT

Primer B: ATGTTCTCTTGGTTCCTCTGAC

STS size: 350

PCR Profile:

Initial incubation: 95 degrees C for 10 minutes

Denaturation: 94 degrees C for 30 seconds

Annealing: 60 degrees C for 30 seconds

Polymerization: 72 degrees C for 23 seconds

PCR Cycles: 30

Thermal Cycler: Perkin Elmer 9700

Protocol:

Template: 25 ng

Primer: each 1 uM

dNTPs: each 200 uM

AmpliTaq Gold Polymerase: 0.07 units/ul

Total Vol: 5 ul

Buffer:

MgCl2: 2.5 mM

KCl: 50 mM

Tris-HCl: 10 mM

pH: 8.3

Cosmid end sequences from generated from flow-sorted chromosome 4 (Human Genome Center, Los Alamos National Laboratory), sequenced at the Stanford Human Genome Center.

FEATURES

source 1..761

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/map="4"

/clone\_lib="Human"

STS 220..569

primer\_bind 220..569

primer\_bind complement(547..569)

BASE COUNT 135 a 235 c 204 g 187 t

ORIGIN

Query Match 100.0%; Score 12; DB 11; Length 761;

Best Local Similarity 100.0%; Pred. No. 4.4e+03;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12

|||||

Db 298 CAGGAGCCTCGA 287

RESULT 15

A70150/c

LOCUS A70150 822 bp DNA linear PAT 07-MAY-1999

DEFINITION Sequence 11 from Patent WO9811228.

ACCESSION A70150

VERSION A70150.1 GI:4774565

KEYWORDS .

SOURCE unidentified.

ORGANISM unidentified.

REFERENCE 1 (bases 1 to 822)

AUTHORS Seymour, G.B., Bird, C.R. and Medina-Suarez, R.D.

TITLE GENETIC CONTROL OF FRUIT RIPENING

JOURNAL Patent: WO 9811228-A 11 19-MAR-1998;

SEYMOUR GRAHAM BARRON (GB)

FEATURES

source 1..822

/organism="unidentified"

/db\_xref="taxon:32644"

/clone="U-055"

BASE COUNT 145 a 254 c 224 g 159 t

ORIGIN

Query Match 100.0%; Score 12; DB 6; Length 822;

Best Local Similarity 100.0%; Pred. No. 4.4e+03;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12

|||||

Db 310 CAGGAGCCTCGA 299

Search completed: July 31, 2002, 14:01:55

Job time: 17557 sec

---





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:21:09 ; Search time 165.21 Seconds  
(without alignments)  
38.657 Million cell updates/sec

Title: US-09-899-718a-5

Perfect score: 26

Sequence:

1 cccgtctaggcgttcgggtgcggcc 26

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*

3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*

4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*

5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length  | DB ID | Description        |
|------------|-------|-------------|---------|-------|--------------------|
| C 1        | 18.6  | 71.5        | 1248    | 4     | US-09-105-537-7    |
| C 2        | 18.6  | 71.5        | 13613   | 4     | US-09-105-537-3    |
| C 3        | 16.6  | 63.8        | 3061    | 1     | US-09-570-842-1    |
| C 4        | 16.6  | 63.8        | 3061    | 1     | US-08-700-576-1    |
| C 5        | 16.6  | 63.8        | 15664   | 1     | US-08-402-282-3    |
| C 6        | 16.6  | 63.8        | 15664   | 1     | US-08-508-004-3    |
| C 7        | 16.6  | 63.8        | 15664   | 1     | US-08-402-066-3    |
| C 8        | 16.6  | 63.8        | 15664   | 1     | US-08-402-088-3    |
| C 9        | 16.6  | 63.8        | 43280   | 2     | US-08-804-227C-1   |
| C 10       | 16.4  | 63.1        | 4403765 | 4     | US-09-103-840A-2   |
| C 11       | 16.4  | 63.1        | 4411529 | 4     | US-09-103-840A-1   |
| C 12       | 16    | 61.5        | 624     | 1     | US-07-661-610C-14  |
| C 13       | 16    | 61.5        | 1008    | 4     | US-09-199-637A-148 |
| C 14       | 16    | 61.5        | 1392    | 4     | US-08-957-351-1    |
| C 15       | 16    | 61.5        | 2101    | 2     | US-08-637-763B-5   |
| C 16       | 16    | 61.5        | 2101    | 3     | US-09-170-354-5    |
| C 17       | 16    | 61.5        | 5437    | 1     | US-07-661-610C-1   |
| C 18       | 16    | 61.5        | 44377   | 2     | US-08-804-227C-7   |
| C 19       | 16    | 61.5        | 44377   | 2     | US-08-804-198-1    |
| C 20       | 16    | 61.5        | 80161   | 3     | US-09-036-987A-1   |
| C 21       | 16    | 61.5        | 80161   | 4     | US-09-370-700-1    |
| C 22       | 15.8  | 60.8        | 68750   | 3     | US-09-335-409-1    |
| C 23       | 15.8  | 60.8        | 68750   | 4     | US-09-568-102-1    |
| C 24       | 15.8  | 60.8        | 68750   | 4     | US-09-567-969-1    |
| C 25       | 15.8  | 60.8        | 68750   | 4     | US-09-568-480-1    |
| C 26       | 15.8  | 60.8        | 68750   | 4     | US-09-568-486-1    |
| C 27       | 15.8  | 60.8        | 68750   | 4     | US-09-568-472-1    |

```

c 28 15.8 60.8 71989 4 US-09-443-501A-2 Sequence 2, Appli
c 29 15.6 60.0 1941 4 US-09-232-200-28 Sequence 28, Appli
c 30 15.6 60.0 1941 4 US-09-232-197-28 Sequence 28, Appli
c 31 15.6 60.0 1941 4 US-09-232-201-28 Sequence 28, Appli
c 32 15.6 60.0 2220 4 US-08-765-907A-14 Sequence 14, Appli
c 33 15.6 60.0 3694 4 US-09-232-197-46 Sequence 46, Appli
c 34 15.6 60.0 3694 4 US-09-232-197-46 Sequence 46, Appli
c 35 15.6 60.0 3694 4 US-09-232-201-46 Sequence 46, Appli
c 36 15.6 60.0 3704 4 US-09-232-200-24 Sequence 24, Appli
c 37 15.6 60.0 3704 4 US-09-232-197-24 Sequence 24, Appli
c 38 15.6 60.0 3704 4 US-09-232-201-24 Sequence 24, Appli
c 39 15.6 60.0 4496 4 US-08-765-907A-6 Sequence 6, Appli
c 40 15.6 60.0 18318 1 US-08-414-926A-6 Sequence 6, Appli
c 41 15.6 60.0 18318 2 US-08-926-922-6 Sequence 6, Appli
c 42 15.6 60.0 18318 3 US-09-253-682-6 Sequence 6, Appli
c 43 15.6 60.0 18318 4 US-09-527-657-6 Sequence 6, Appli
c 44 15.4 59.2 960 1 US-07-639-330E-1 Sequence 1, Appli
c 45 15.4 59.2 960 3 US-09-248-335-57 Sequence 57, Appli

```

## ALIGNMENTS

```

RESULT 1
US-09-105-537-7/c
; Sequence 7, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: fastseq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-7

```

```

Query Match 71.5%; Score 18.6; DB 4; Length 1248;
Best Local Similarity 84.0%; Pred. No. 5;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Oy 1 cccgtctaggcgttcgggtgcggc 25
Db 1068 CCGCGCGAGCGGTCGGTGGCGGC 1044

```

```

RESULT 2
US-09-105-537-3
; Sequence 3, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae

```



```

; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 747..1109
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1109..2014
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2034..2747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2747..3109
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3109..3444
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3444..3728
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3731..4855
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4855..5376
; OTHER INFORMATION: /function= "potential coding
; OTHER INFORMATION: sequence"
; OTHER INFORMATION: /product= "L5 gp37 homolog"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5382..5747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5837..6307
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6403..7770
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7770..8006
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8033..8236
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8244..9443
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:

```

```

; NAME/KEY: misc_feature
; LOCATION: 9450..10244
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10371..10586
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11115..11786
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11917..12741
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 12748..14499
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14771..15154
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15154..15426
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15429..15664
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; US-08-402-282-3

```

Query Match 63.8%; Score 16.6; DB 1; Length 15664;  
Best Local Similarity 82.6%; Pred. No. 44;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 cgtctaggcgttcggtccggc 25  
DB 11139 CGTCCAGGATTCGGTGTCCGTC 11117

RESULT 6  
US-08-508-004-3/c  
; Sequence 3, Application US/08508004  
; Patent No. 5582969  
; GENERAL INFORMATION:  
; APPLICANT: Pearson, Robert E.  
; APPLICANT: Dickson, Julie A.  
; APPLICANT: Hamilton, Paul T.  
; APPLICANT: Little, Michael C.  
; APPLICANT: Beyer Jr., Wayne F.  
; TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Richard J. Rodrick, Becton Dickinson and  
; ADDRESSEE: Company  
; STREET: 1 Becton Drive  
; CITY: Franklin Lakes  
; STATE: NJ  
; COUNTRY: US  
; ZIP: 07417  
; COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,004
FILING DATE: 27-JUL-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/402,282
FILING DATE: 10-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fugit, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-3283
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 15664 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 222..425
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 451..747
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 747..1109
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1109..2014
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2034..2747
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2747..3109
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3109..3444
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3444..3728
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3731..4855
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4855..5376
OTHER INFORMATION: /function= "potential coding
OTHER INFORMATION: sequence"
OTHER INFORMATION: /product= "L5 gp37 homolog"

```

Query Match 63.8%; Score 16.6; DB 1; Length 15664;  
Best Local Similarity 82.6%; Pred. No. 44;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 cgtctaggcgttcggttcgcgcg 25  
||||| ||| ||||| ||||| |||  
Db 11139 GCTCAGGATTGCTGCTCGTC 11117

## RESULT 7

US-08-402-066-3/C

; Sequence 3, Application US/08402066

; Patent No. 5612182

; GENERAL INFORMATION:

; APPLICANT: Pearson, Robert E.

; APPLICANT: Dickson, Julie A.

; APPLICANT: Hamilton, Paul T.

; APPLICANT: Little, Michael C.

; APPLICANT: Beyer Jr., Wayne F.

; TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE

; TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Richard J. Rodrick, Becton Dickinson and

; ADDRESSEE: Company

; STREET: 1 Becton Drive

; CITY: Franklin Lakes

; STATE: NJ

; COUNTRY: US

; ZIP: 07417

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/402,066

; FILING DATE:

; CLASSIFICATION: 436

; ATTORNEY/AGENT INFORMATION:

; NAME: Fugit, Donna R.

; REGISTRATION NUMBER: 32,135

; REFERENCE/DOCKET NUMBER: P-3283

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 15664 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: 222..425

; OTHER INFORMATION: /function= "potential open reading

; OTHER INFORMATION: frame"

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: 451..747

; OTHER INFORMATION: /function= "potential open reading

; OTHER INFORMATION: frame"

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: 747..1109

; OTHER INFORMATION: /function= "potential open reading

; OTHER INFORMATION: frame"

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: 1109..2014

; OTHER INFORMATION: /function= "potential open reading

; OTHER INFORMATION: frame"

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: 1115..11786

; OTHER INFORMATION: /function= "potential open reading

; OTHER INFORMATION: frame"

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: 2034..2747

OTHER INFORMATION: /function= "potential open reading  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2747..3109  
OTHER INFORMATION: /function= "potential open reading  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 3109..3444  
OTHER INFORMATION: /function= "potential open reading  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 3444..3728  
OTHER INFORMATION: /function= "potential open reading  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 3731..4855  
OTHER INFORMATION: /function= "potential open reading  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 4855..5376  
OTHER INFORMATION: /function= "potential coding  
OTHER INFORMATION: sequence"  
OTHER INFORMATION: /product= "L5 gp37 homolog"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 5382..5747  
OTHER INFORMATION: /function= "potential open reading  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 5837..6307  
OTHER INFORMATION: /function= "potential open reading  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 6403..7770  
OTHER INFORMATION: /function= "potential open reading  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 7770..8006  
OTHER INFORMATION: /function= "potential open reading  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 8033..8236  
OTHER INFORMATION: /function= "potential open reading  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 8244..9443  
OTHER INFORMATION: /function= "potential open reading  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 9450..10244  
OTHER INFORMATION: /function= "potential open reading  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 10371..10586  
OTHER INFORMATION: /function= "potential open reading  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 11115..11786  
OTHER INFORMATION: /function= "potential open reading  
OTHER INFORMATION: frame"

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 11917..12741  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 12748..14499  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 14771..15134  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 15154..15426  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 15429..15664  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
US-08-402-066-3

Query Match 63.8%; Score 16.6; DB 1; Length 15664;  
Best Local Similarity 82.6%; Pred. No. 44;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 3 cgtctaggctgcgtgtccgc 25  
||||| ||| ||||| |||||  
Db 11139 CGTCCAGGATTGGTGGTCGTC 11117

RESULT 8  
US-08-402-068-3/c  
Sequence 3, Application US/08402068  
Patent No. 5633159  
GENERAL INFORMATION:  
APPLICANT: Pearson, Robert E.  
APPLICANT: Dickson, Julie A.  
APPLICANT: Hamilton, Paul T.  
APPLICANT: Little, Michael C.  
APPLICANT: Beyer Jr., Wayne F.  
TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE  
TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Richard J. Rodrick, Becton Dickinson and  
ADDRESSEE: Company  
STREET: 1 Becton Drive  
CITY: Franklin Lakes  
STATE: NJ  
COUNTRY: US  
ZIP: 07417  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/402,068  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fugit, Donna R.  
REGISTRATION NUMBER: 32,135  
REFERENCE/DOCKET NUMBER: P-3283  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:

LENGTH: 15664 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 222..425  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 451..747  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 747..1109  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1109..2014  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2034..2747  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2747..3109  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 3109..3444  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 3444..3728  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 3731..4855  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 4855..5376  
OTHER INFORMATION: /function= "potential coding  
OTHER INFORMATION: sequence"  
OTHER INFORMATION: /product= "L5 gp37 homolog"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 5382..5747  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 5837..6307  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 6403..7770  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature

```
; LOCATION: 7770..8006
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8033..8236
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8244..9443
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9450..10244
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10371..10586
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11115..11786
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11917..12741
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 12748..14499
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14771..15154
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15154..15426
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15429..15664
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; US-08-402-068-3
```

```
Query Match 63.8%; Score 16.6; DB 1; Length 15664;
Best Local Similarity 82.6%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 3 cgtctaggcgttcggtgcgcgc 25
    ||| ||| ||| ||| ||| ||| |||
Db 11139 CGTCCAGGATTTCGGTGTCCGTC 11117
```

```
RESULT 9
US-08-804-227C-1/c
; Sequence 1, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rosteck, Paul R., Jr.
```

```
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43280 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 816..14234
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14351..19945
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20010..31199
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31232..36067
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36249..41774
; US-08-804-227C-1
```

```
Query Match 63.8%; Score 16.6; DB 2; Length 43280;
Best Local Similarity 82.6%; Pred. No. 46;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 3 cgtctaggcgttcggtgcgcgc 25
    ||| ||| ||| ||| ||| ||| |||
Db 27009 CGTCCAGCGCTCGGTGTCCAGC 26987
```

```
RESULT 10
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
```





RESULT 14  
US-08-957-351-1/c  
; Sequence 1, Application US/08957351  
; Patent No. 6306586  
; GENERAL INFORMATION:  
; APPLICANT: Semina, Elena  
; APPLICANT: Murray, Jeffrey C.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF CATARACTS  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109-2170  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/957,351  
; FILING DATE: 24-OCT-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arnold, Beth E.  
; REGISTRATION NUMBER: 35,430  
; REFERENCE/DOCKET NUMBER: UTA-024.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-832-1000  
; TELEFAX: 617-832-7000  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1392 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-957-351-1

Query Match 61.5%; Score 16; DB 4; Length 1392;  
Best Local Similarity 79.2%; Pred. No. 74;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 cccgtctaggcggttcggtgcgg 24  
||| ||||| ||| |||||  
Db 1119 CCACTCTAGCGACCCCTGTCCG 1096

RESULT 15  
US-08-637-763B-5  
; Sequence 5, Application US/08637763B  
; Patent No. 5849559  
; GENERAL INFORMATION:  
; APPLICANT: VAN DER WOUW, Monique J.A. et al  
; TITLE OF INVENTION: ARABINOXYLAN DEGRADING ENZYME  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morrison & Foerster  
; STREET: 2000 Pennsylvania Avenue, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/637,763B  
; FILING DATE: 25-AUG-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murashige, Kate H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 4615-0066.00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030 MRSNFOERSWSH  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2101 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Aspergillus niger  
; STRAIN: CBS 120.49  
; FEATURE:  
; NAME/KEY: TATA\_signal  
; LOCATION: 665..670  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 784..1779  
; OTHER INFORMATION: /product= "arabinoxylan degrading  
; OTHER INFORMATION: /enzyme= "axda"  
; OTHER INFORMATION: /gene= "axda"  
; OTHER INFORMATION: /standard\_name= "arabinoxylan degrading enzyme"  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 784..861  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 862..1779  
US-08-637-763B-5

Query Match 61.5%; Score 16; DB 2; Length 2101;  
Best Local Similarity 79.2%; Pred. No. 75;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 cccgtctaggcggttcggtgcgg 24  
||| ||||| ||| |||||  
Db 266 CACGTCTAGACTTTCGATGCCCGG 289

Search completed: July 31, 2002, 12:23:47  
Job time: 11850 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:12:01 ; Search time 5855.71 Seconds  
(without alignments)  
59.928 Million cell updates/sec

Title: US-09-899-718A-5  
Perfect score: 26  
Sequence: 1 cccgtctaggcttcgggttcggcc 26

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_esttc:\*  
9: gb\_estli:\*  
10: gb\_estt2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pin:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID       | Description        |
|------------|-------|-------------|--------|----|----------|--------------------|
| 1          | 19.2  | 73.8        | 668    | 12 | BH506552 | BH506552 BOGR94TF  |
| 2          | 19.2  | 73.8        | 747    | 12 | BH428842 | BH428842 BOGR94TF  |
| 3          | 18.6  | 71.5        | 420    | 12 | BH391002 | BH391002 AG-ND-180 |
| 4          | 18.6  | 71.5        | 700    | 10 | BM072919 | BM072919 MEST55-H0 |
| 5          | 18.2  | 70.0        | 576    | 10 | BF770774 | BF770774 RCI-IT001 |
| 6          | 18.2  | 70.0        | 771    | 12 | CNS02PLG | AL208141 Tetraodon |
| 7          | 18.6  | 69.2        | 319    | 12 | A2496643 | A2496643 IM0333K08 |
| 8          | 18.6  | 69.2        | 745    | 10 | BM398961 | BM398961 5009-0-51 |
| 9          | 18.6  | 69.2        | 845    | 10 | BE748192 | BE748192 601571585 |
| 10         | 18.6  | 69.2        | 1031   | 12 | AG060278 | AG060278 Pan trogl |
| 11         | 18.6  | 69.2        | 1425   | 10 | BE869446 | BE869446 601446101 |
| 12         | 17.8  | 68.5        | 633    | 9  | AW949909 | AW949909 EST361979 |
| 13         | 17.6  | 67.7        | 495    | 10 | BG818796 | BG818796 602779143 |
| 14         | 17.6  | 67.7        | 499    | 10 | BE462300 | BE462300 EST324564 |
| 15         | 17.6  | 67.7        | 591    | 10 | BI490136 | BI490136 603031972 |
| 16         | 17.6  | 67.7        | 652    | 10 | BI925736 | BI925736 EST545625 |
| 17         | 17.6  | 67.7        | 702    | 10 | BF867684 | BF867684 963093D05 |

|      |      |      |      |    |          |                    |
|------|------|------|------|----|----------|--------------------|
| C 18 | 17.6 | 67.7 | 824  | 10 | BE892849 | BE892849 601433981 |
| C 19 | 17.6 | 67.7 | 867  | 10 | BF241175 | BF241175 601880044 |
| C 20 | 17.6 | 67.7 | 877  | 10 | BI462741 | BI462741 603202191 |
| C 21 | 17.6 | 67.7 | 911  | 10 | BI601214 | BI601214 603245726 |
| C 22 | 17.6 | 67.7 | 932  | 10 | BF180488 | BF180488 601805280 |
| C 23 | 17.6 | 67.7 | 1147 | 10 | BI834829 | BI834829 603090281 |
| C 24 | 17.2 | 66.2 | 215  | 10 | BF770115 | BF770115 RCI-IT001 |
| C 25 | 17.2 | 66.2 | 223  | 10 | BE551977 | BE551977 hy02b02.x |
| C 26 | 17.2 | 66.2 | 314  | 9  | AL133850 | AL133850 DKFp761F  |
| C 27 | 17.2 | 66.2 | 348  | 10 | BE937644 | BE937644 MRI-TN002 |
| C 28 | 17.2 | 66.2 | 348  | 10 | BE937659 | BE937659 MRI-TN002 |
| C 29 | 17.2 | 66.2 | 355  | 10 | BE937624 | BE937624 MRI-TN002 |
| C 30 | 17.2 | 66.2 | 376  | 10 | N60925   | N60925 TgESTzyl4d0 |
| C 31 | 17.2 | 66.2 | 421  | 10 | H20321   | H20321 ym62q01.s1  |
| C 32 | 17.2 | 66.2 | 438  | 10 | H15755   | H15755 y127d10.r1  |
| C 33 | 17.2 | 66.2 | 458  | 9  | AI222008 | AI222008 qg95d11.x |
| C 34 | 17.2 | 66.2 | 520  | 9  | BE161596 | BE161596 MR0-HT044 |
| C 35 | 17.2 | 66.2 | 539  | 10 | BI774972 | BI774972 467308 MA |
| C 36 | 17.2 | 66.2 | 603  | 10 | BG992087 | BG992087 MR2-HT116 |
| C 37 | 17.2 | 66.2 | 607  | 10 | BG982963 | BG982963 PM0-CN015 |
| C 38 | 17.2 | 66.2 | 610  | 9  | AI398765 | AI398765 NCW09A777 |
| C 39 | 17.2 | 66.2 | 665  | 10 | BG983477 | BG983477 PM0-CN015 |
| C 40 | 17.2 | 66.2 | 681  | 10 | BG419972 | BG419972 602453646 |
| C 41 | 17.2 | 66.2 | 707  | 10 | BG479901 | BG479901 602705750 |
| C 42 | 17.2 | 66.2 | 733  | 10 | BG428650 | BG428650 602494788 |
| C 43 | 17.2 | 66.2 | 756  | 10 | BE727948 | BE727948 601560594 |
| C 44 | 17.2 | 66.2 | 818  | 10 | BF316519 | BF316519 601903032 |
| C 45 | 17.2 | 66.2 | 847  | 10 | BG281333 | BG281333 602401814 |

## ALIGNMENTS

RESULT 1  
BH506552  
LOCUS  
DEFINITION  
BOGR94TF BOGF Brassica oleracea genomic clone BOGR94, DNA  
668 bp  
linear  
GSS 13-DEC-2001  
sequence.  
ACCESSION  
BH506552  
VERSION  
BH506552.1  
KEYWORDS  
GSS.  
SOURCE  
Brassica oleracea.  
ORGANISM  
Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE  
1 (bases 1 to 668)  
Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
Other GSSs: BOGR94TF  
Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TF  
Class: sheared ends.  
Location/Qualifiers  
1.668  
/organism="Brassica oleracea"  
/strain="TOL000H3"  
/db\_xref="taxon:3712"  
/clone="BOGR94"  
/note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

## FEATURES

source  
195 a 140 c 125 g 208 t  
BASE COUNT  
ORIGIN

Query Match 73.8%; Score 19.2; DB 12; Length 668;  
 Best Local Similarity 87.5%; Pred. No. 3.8e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ccgtctaggcgttcggttcggc 25  
 Db 534 CAGTCTTCGGTTCGGTTCGGC 557  
 1 ||||| ||||| ||||| ||||| |||||

RESULT 2  
 BH428842 747 bp DNA linear GSS 12-DEC-2001  
 LOCUS BOGS283TR BOGS Brassica oleracea genomic clone BOGS283, DNA  
 DEFINITION sequence.  
 ACCESSION BH428842  
 VERSION BH428842.1 GI:17614563  
 KEYWORDS GSS.  
 SOURCE Brassica oleracea.  
 ORGANISM Brassica oleracea  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 747)  
 TOWN, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.  
 Whole genome shotgun sequencing of Brassica oleracea  
 JOURNAL Unpublished (2001)  
 COMMENT Other\_GSSs: BOGS283TF  
 Contact: Chris Town  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.  
 Seq primer: TR  
 Class: sheared ends.

FEATURES  
 source  
 1..747  
 /organism="Brassica oleracea"  
 /strain="T01000DH3"  
 /db\_xref="taxon:3712"  
 /clone\_lib="BOGS283"  
 /clone\_lib="BOGS"  
 /note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared  
 genomic DNA inserted into pHOS1 using BstXI linkers"

BASE COUNT 204 a 150 c 128 g 265 t

Query Match 73.8%; Score 19.2; DB 12; Length 747;  
 Best Local Similarity 87.5%; Pred. No. 3.9e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ccgtctaggcgttcggttcggc 25  
 Db 454 CAGTCTTCGGTTCGGTTCGGC 477  
 1 ||||| ||||| ||||| ||||| |||||

RESULT 3  
 BH391002/2 420 bp DNA linear GSS 11-DEC-2001  
 LOCUS AG-ND-180A21, TR ND-TAM Anopheles gambiae genomic clone AG-ND-180A21  
 DEFINITION , DNA sequence.  
 ACCESSION BH391002  
 VERSION BH391002.1 GI:17337143  
 KEYWORDS GSS.  
 SOURCE African malaria mosquito.  
 ORGANISM Anopheles gambiae  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae  
 ; Anopheles.  
 REFERENCE 1 (bases 1 to 420)  
 SHETTY, J., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B.J.

TITLE Direct Submission of BAC-end sequences from Anopheles gambiae  
 JOURNAL Unpublished (2001)  
 COMMENT Other\_GSSs: AG-ND-180A21.TF  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: bjlftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.  
 Seq primer: M13 Rev  
 Class: BAC ends.

FEATURES  
 source  
 1..420  
 /organism="Anopheles gambiae"  
 /strain="PEST"  
 /db\_xref="taxon:7165"  
 /clone="AG-ND-180A21"  
 /clone\_lib="ND-TAM"  
 /note="Vector: pECBAC1; Site\_1: HindIII"

BASE COUNT 75 a 135 c 120 g 90 t

Query Match 71.5%; Score 18.6; DB 12; Length 420;  
 Best Local Similarity 84.0%; Pred. No. 6.2e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ccgtctaggcgttcggttcggc 25  
 Db 62 CCCTCTCGCGTTCGGTTCGGC 38  
 1 ||||| ||||| ||||| ||||| |||||

RESULT 4

BM072919 700 bp mRNA linear EST 13-NOV-2001  
 LOCUS MEST55-H08.T3 ISUM4-TN Zea mays cDNA clone MEST55-H08 3', mRNA  
 DEFINITION sequence.

ACCESSION BM072919  
 VERSION BM072919.1 GI:16916614  
 KEYWORDS EST.  
 SOURCE Zea mays.

ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 700)  
 QIU, F., CUI, F., GUO, L., ASHLOCK, D.A., WEN, F.J. and SCHNABLE, P.S.  
 Expressed Sequence Tags from B73 Maize Seedlings and Silks  
 JOURNAL Unpublished (2001)

AUTHORS Contact: Patrick S. Schnable  
 TITLE Schnable Laboratory  
 JOURNAL Iowa State University  
 COMMENT G405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA  
 Tel: 515-294-0975  
 Fax: 515-294-2299  
 Email: schnable@iastate.edu

Individual basecall and confidence value were assigned using the Phred software,  
 (<http://depts.washington.edu/ventures/collabtr/direct/index.htm#b  
 rt>). Overall sequence quality assessment and vector trimming were  
 conducted using the Lucy software (<http://www.tigr.org/softlab/>).  
 Lucy parameters were set to ensure an overall trimmed quality of  
 97.5% or better without any vector fragments in the chosen

high-quality region of each sequence. Low-quality bases between the poly-T and the high-quality region were replaced with N's to serve as spacers.

PCR Primers  
 FORWARD: Forward PCR primer sequence, primer T7-1 (AA TAC GAC TCA CTA TAG)  
 BACKWARD: Backward PCR primer sequence, primer T3 (ATT AAC CCT CAC TAA AG)  
 TAA AG)

Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).

#### FEATURES

Location/Qualifiers

1. 700

/organism="Zea mays"

/cultivar="B73"

/db\_xref="taxon:4577"

/clone="WEST55-H08"

/clone\_lib="ISUM4-TN"

/tissue\_type="Seedling and silk"

/lab\_host="DH108"

/note="Vector: pT7T3PAC; Site\_1: EcoRI; Site\_2: NotI; ds-cDNA molecules were generated as follows. First-strand cDNA was prepared from oligo-dT selected mRNA by priming with a NotI oligo-dT primer (5' AACTGGAAGATTCGGCGCCGAGGAATTTTTTTTTTTTTT). The resulting DNA:RNA hybrid was treated with RNase H and used as a template for DNA PolI-catalyzed second strand synthesis. After the addition of EcoRI adaptors, the ds-cDNAs were digested with NotI and size-selected. The resulting molecules were directionally cloned into the EcoRI and NotI sites of the pT7T3PAC vector. The library then went through one round of normalization to Cot value of 5 based on the methods of Marcelo Bento Soares (Genome Research 6: 791-806, 1996)."

BASE COUNT 173 a 153 c 167 g 207 t  
 ORIGIN

Query Match 71.5%; Score 18.6; DB 10; Length 700;

Best Local Similarity 84.0%; Pred. No. 6.8e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ccgcctagcgctcggtgcggc 25

Db 299 CCCATCTCGCGCTCGTGTCTCGTCG 323

#### RESULT

5

LOCUS

BF770774

DEFINITION RC1-IT0013-251100-012-e09 IT0013 Homo sapiens cDNA, mRNA sequence.

ACCESSION BF770774

VERSION BF770774.1 GI:12118674

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 576)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,

Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

#### TITLE

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC1st2-RC1-IT0013-251100-012-e09st3-2000-11-25&tl=1)

Seq primer: puc 18 forward

High quality sequence start: 31

High quality sequence stop: 575.

#### FEATURES

source

1. 576

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="IT0013"

/dev\_stage="Adult"

/note="Organ: epid\_tumor; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (O.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 109 a 142 c 206 g 119 t

#### ORIGIN

Query Match 70.0%; Score 18.2; DB 10; Length 576;

Best Local Similarity 87.0%; Pred. No. 9.6e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 cgcctagcgctcggtgcggc 25

Db 508 CGCTAGCGCTCTGTCTCAGGC 530

#### RESULT

6

CNS02PLG

LOCUS

DEFINITION

771 bp DNA linear GSS 14-MAY-2000

Tetraodon nigroviridis genome survey sequence T7 end of clone

156C05 of library G from Tetraodon nigroviridis, genomic survey

sequence.

AL208141

AL208141.1 GI:7866960

GSS: genome survey sequence.

Tetraodon nigroviridis.

Tetraodon nigroviridis.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;

Tetraodontidae; Tetraodon.

1 (bases 1 to 771)

Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,

Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and

Weissenbach,J.

Characterization and repeat analysis of the compact genome of the

freshwater pufferfish Tetraodon nigroviridis

Unpublished

2 (bases 1 to 771)

Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,

Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,

Saurin,W. and Weissenbach,J.

Human gene number estimate provided by genome wide analysis using

Tetraodon nigroviridis DNA sequence

Unpublished

3 (bases 1 to 771)

Genoscope.

Direct Submission

Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

This sequence is a single read and was generated as part of a large

scale clone-end sequencing project of the Tetraodon nigroviridis

genome. For more information, please take a look at

http://www.genoscope.cns.fr/Tetraodon.

#### FEATURES

Location/Qualifiers

source 1. .771  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="156C05"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COAG156AB03LP1-end : T7"  
BASE COUNT 116 a 236 c 267 g 139 t 13 others  
ORIGIN

Query Match 70.0%; Score 18.2; DB 12; Length 771;  
Best Local Similarity 87.0%; Pred. No. 1e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 3 cgtctagcgttcggtggtccggc 25  
||||| ||||| ||||| ||||| |||||  
Db 570 GGTCCAGCGCGGTGGTGTCCGCG 592

RESULT 7  
A2496643/c 319 bp DNA linear GSS 05-OCT-2000  
LOCUS 1M0333K08F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
DEFINITION clone UUGC1M0333K08 F, DNA sequence.  
ACCESSION A2496643  
VERSION A2496643.1 GI:10672981  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 319)  
AUTHORS Dunn,D., Ayagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A., and Wright,D., Weiss,R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0333 row: K column: 08  
Seq primer: CGTTGTAAACGACGCGCCAGT  
Class: plasmid ends  
High quality sequence stop: 319.  
FEATURES  
source Location/Qualifiers  
1. .319  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0333K08"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid RL. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance." 53 t 1 others  
BASE COUNT 44 a 105 c 116 g 53 t 1 others  
ORIGIN

Query Match 69.2%; Score 18; DB 12; Length 319;  
Best Local Similarity 80.8%; Pred. No. 1e+03;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 cccgtctagcgttcggtggtccggc 26  
||||| ||||| ||||| ||||| |||||  
Db 187 CCGTCCGCGCGCGGTCTCCGCGC 162

RESULT 8  
BM398961/c 745 bp mRNA linear EST 17-JAN-2002  
LOCUS 5009-0-51-B07.t.1 Chilcoat/Turkewitz cDNA (large fraction)  
DEFINITION Tetrahymena thermophila cDNA, mRNA sequence.  
ACCESSION BM398961  
VERSION BM398961.1 GI:18199014  
KEYWORDS EST.  
SOURCE Tetrahymena thermophila.  
ORGANISM Tetrahymena thermophila  
REFERENCE 1 (bases 1 to 745)  
AUTHORS Turkewitz,A.P., Karrer,K.M., Jahn,C., Orias,E., Kirk,K.E., Frankel,J., and Klobutcher,L.  
TITLE EST from Tetrahymena thermophila, strain CU428.1, growing cells  
JOURNAL Unpublished (2002)  
COMMENT Contact: Turkewitz AP  
Molecular Genetics and Cell Biology  
University of Chicago  
920 E. 58th Street, Chicago, IL 60637, USA  
Tel: 773 702 4374  
Fax: 773 702 3172  
Email: apturkew@midway.uchicago.edu  
Seq primer: T3.  
FEATURES  
source Location/Qualifiers  
1. .745  
/organism="Tetrahymena thermophila"  
/strain="CU428.1"  
/db\_xref="taxon:5911"  
/clone\_lib="Chilcoat/Turkewitz cDNA (large fraction)"  
/note="Vector: Bluescript2 SK+; Details on library preparation can be found in Chilcoat and Turkewitz (2001) Proc. Natl. Acad. Sci USA, 98: 8709-8713."

BASE COUNT 274 a 146 c 130 g 195 t  
ORIGIN

Query Match 69.2%; Score 18; DB 10; Length 745;  
Best Local Similarity 80.8%; Pred. No. 1.2e+03;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 cccgtctagcgttcggtggtccggc 26  
||||| ||||| ||||| ||||| |||||  
Db 672 CCGTTCGCGCGGTCTGGTCTCCGCGC 647

RESULT 9  
BE748192/c 845 bp mRNA linear EST 15-SEP-2000  
LOCUS 601571585F1 NIH\_MGC\_55 Homo sapiens cDNA clone IMAGE:3838557 5',  
DEFINITION mRNA sequence.  
ACCESSION BE748192

```

VERSION BE748192.1 GI:10162184
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL 1 (bases 1 to 845)
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONETECH Laboratories, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM525 row: i column: 22
High quality sequence stop: 558.
FEATURES
source
Location/Qualifiers
1..845
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3838557"
/clone_lib="NIH_MGC_55"
/tissue_type="from acute myelogenous leukemia"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
Site_1: SfiI (ggccgctcgcc); Site_2: SfiI (ggccataggcc
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCATTAGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGCGCGGACATG-3' (30)BN-3'
(where B = A, C, G and N = A, C, G, or T). Average
insert size 1.65 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
1 others
BASE COUNT 180 a 260 c 232 g 172 t
ORIGIN

Query Match 69.2%; Score 18; DB 10; Length 845;
Best Local Similarity 80.8%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ccgcgttagcggtcggtcggtcgcc 26
||||| ||||| ||||| ||||| |||||
Db 673 CCGGTCAAGCGCGTCGTCGCCGCC 648

RESULT 10
AG060278
LOCUS 1031 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-047L15.F, genomic survey sequence.
ACCESSION AG060278
VERSION AG060278.1 GI:16611510
KEYWORDS GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC Library clone:PTB-047L15.F.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Pan.
1 (sites)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
BAC end sequences of Library PTB
Unpublished
TITLE BAC end sequences of Library PTB
JOURNAL
REFERENCE 2 (bases 1 to 1031)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.

Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsuri-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimps@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1..1031
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-047L15.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 103 a 227 c 400 g 262 t
ORIGIN 39 others

Query Match 69.2%; Score 18; DB 12; Length 1031;
Best Local Similarity 80.8%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ccgcgttagcggtcggtcggtcgcc 26
||||| ||||| ||||| ||||| |||||
Db 392 CCGGCCAGCGCGCGGTCGCCGCC 417

RESULT 11
BE869446/c
LOCUS 1425 bp mRNA linear EST 20-OCT-2000
DEFINITION 601446101F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3849888 5',
mRNA sequence.
ACCESSION BE869446
VERSION BE869446.1 GI:10318222
KEYWORDS EST.
SOURCE human.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1425)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9568 row: b column: 01
High quality sequence stop: 154.
Location/Qualifiers
1..1425
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3849888"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
FEATURES
source

```









GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:11:17 ; Search time 720.8 Seconds

(without alignments)  
61.931 Million cell updates/sec

Title: US-09-899-718A-5

Perfect score: 26

Sequence: 1 ccgcctaggcgttcggtccgccc 26

Scoring table: IDENTIFY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_032802.\*

- 1: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1980.DAT.\*
- 2: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1981.DAT.\*
- 3: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1982.DAT.\*
- 4: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1983.DAT.\*
- 5: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1984.DAT.\*
- 6: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1985.DAT.\*
- 7: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1986.DAT.\*
- 8: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1987.DAT.\*
- 9: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1988.DAT.\*
- 10: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1989.DAT.\*
- 11: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1990.DAT.\*
- 12: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1991.DAT.\*
- 13: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1992.DAT.\*
- 14: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1993.DAT.\*
- 15: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1994.DAT.\*
- 16: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1995.DAT.\*
- 17: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1996.DAT.\*
- 18: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1997.DAT.\*
- 19: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1998.DAT.\*
- 20: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1999.DAT.\*
- 21: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA2000.DAT.\*
- 22: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA2001A.DAT.\*
- 23: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA2001B.DAT.\*
- 24: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| c 1        | 18.6  | 71.5        | 1248   | 21 | AAZ87286    |
| c 2        | 18.6  | 71.5        | 12441  | 21 | AAZ87284    |
| c 3        | 18.6  | 71.5        | 13613  | 21 | AAZ87319    |
| c 4        | 17.6  | 67.7        | 47981  | 22 | AAZ87357    |
| c 5        | 17.2  | 66.2        | 1680   | 22 | AAH41415    |
| c 6        | 17    | 65.4        | 381    | 22 | ABA46067    |
| c 7        | 17    | 65.4        | 381    | 22 | ABA56605    |
| c 8        | 17    | 65.4        | 381    | 22 | ABA26220    |
| c 9        | 17    | 65.4        | 381    | 22 | AAK04749    |

|      |      |      |         |    |           |                      |
|------|------|------|---------|----|-----------|----------------------|
| c 10 | 17   | 65.4 | 381     | 22 | AAK30275  | Human bone marrow    |
| c 11 | 17   | 65.4 | 381     | 22 | AAI14871  | Probe #4804 for ge   |
| c 12 | 17   | 65.4 | 381     | 22 | AAI36227  | Probe #4913 used t   |
| c 13 | 17   | 65.4 | 381     | 22 | AAI04658  | Probe #4649 used t   |
| c 14 | 17   | 65.4 | 1463    | 11 | AAQ06685  | Polygisticronic oper |
| c 15 | 17   | 65.4 | 2443    | 22 | AAH35021  | Human colon cancer   |
| c 16 | 17   | 65.4 | 32249   | 22 | AAH04931  | Human secreted pro   |
| c 17 | 16.6 | 63.8 | 120     | 21 | AAI12097  | Human secreted pro   |
| c 18 | 16.6 | 63.8 | 3061    | 19 | AAV22799  | Nucleotide sequenc   |
| c 19 | 16.6 | 63.8 | 4599    | 20 | AAV57803  | G. oxydans autonom   |
| c 20 | 16.6 | 63.8 | 15664   | 17 | AAAT09312 | Mycobacteriophage    |
| c 21 | 16.6 | 63.8 | 15664   | 18 | AAAT0492  | Mycobacteriophage    |
| c 22 | 16.6 | 63.8 | 15664   | 18 | AAAT66130 | Mycobacteriophage    |
| c 23 | 16.6 | 63.8 | 15664   | 18 | AAAT51224 | NheI-G/SpeI fragme   |
| c 24 | 16.6 | 63.8 | 43280   | 18 | AAAT80413 | Tylosine synthase    |
| c 25 | 16.4 | 63.1 | 952     | 23 | ABL27597  | Drosophila melanog   |
| c 26 | 16.4 | 63.1 | 1175    | 21 | AAF22378  | Human secreted pro   |
| c 27 | 16.4 | 63.1 | 1343    | 22 | ABA16862  | Human nervous syst   |
| c 28 | 16.4 | 63.1 | 1343    | 22 | ABA19759  | Human nervous syst   |
| c 29 | 16.4 | 63.1 | 1571    | 23 | ABL06699  | Drosophila melanog   |
| c 30 | 16.4 | 63.1 | 1959    | 23 | ABL18703  | Drosophila melanog   |
| c 31 | 16.4 | 63.1 | 2310    | 22 | AAAS03788 | Mycobacterium tube   |
| c 32 | 16.4 | 63.1 | 2495    | 12 | AAQ11271  | Prepro-polygalactu   |
| c 33 | 16.4 | 63.1 | 2547    | 22 | AAF93794  | Human cDNA encodin   |
| c 34 | 16.4 | 63.1 | 3090    | 22 | AAF77899  | Quorum sensing con   |
| c 35 | 16.4 | 63.1 | 3442    | 23 | ABL07609  | Drosophila melanog   |
| c 36 | 16.4 | 63.1 | 3728    | 23 | ABL06698  | Drosophila melanog   |
| c 37 | 16.4 | 63.1 | 4109    | 23 | ABL18702  | Drosophila melanog   |
| c 38 | 16.4 | 63.1 | 12713   | 23 | ABL07608  | Drosophila melanog   |
| c 39 | 16.4 | 63.1 | 24221   | 23 | ABL09808  | Drosophila melanog   |
| c 40 | 16.4 | 63.1 | 4403765 | 22 | AAI99683  | Mycobacterium tube   |
| c 41 | 16.4 | 63.1 | 4411529 | 22 | AAI99682  | Mycobacterium tube   |
| c 42 | 16.2 | 62.3 | 303     | 22 | AAK8761   | Human digestive sy   |
| c 43 | 16.2 | 62.3 | 303     | 22 | AAAS31795 | Human liver associ   |
| c 44 | 16.2 | 62.3 | 528     | 22 | AAF30761  | Megalomycin polyke   |
| c 45 | 16.2 | 62.3 | 580     | 23 | ABL25667  | Drosophila melanog   |

## ALIGNMENTS

|        |                                                                         |
|--------|-------------------------------------------------------------------------|
| RESULT | 1                                                                       |
| ID     | AAZ87286/C                                                              |
| XX     | AAZ87286 standard; DNA; 1248 BP.                                        |
| AC     | AAZ87286;                                                               |
| XX     | 05-JUN-2000 (first entry)                                               |
| DT     | S. venezuelae desosamine biosynthetic gene desI, SEQ ID NO:7.           |
| DE     | Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin; |
| DE     | neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;       |
| KW     | biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,    |
| KW     | chronic obstructive pulmonary disease; respiratory inflammation;        |
| KW     | hypercholesterolaemia; crop protection agent; ds.                       |
| XX     | Streptomyces venezuelae ATCC15439.                                      |
| OS     |                                                                         |
| XX     |                                                                         |
| FH     | Key                                                                     |
| FT     | Location/Qualifiers                                                     |
| CDS    | 1..1248                                                                 |
| FT     | /*tag= a                                                                |
| FT     | /product= "DesI"                                                        |
| XX     |                                                                         |
| XX     | WO200000620-A2.                                                         |
| PN     |                                                                         |
| XX     |                                                                         |
| PD     | 06-JAN-2000.                                                            |
| XX     |                                                                         |
| XX     |                                                                         |
| PF     | 25-JUN-1999; 99WO-US14398.                                              |
| XX     |                                                                         |
| XX     |                                                                         |
| PR     | 26-JUN-1998; 98US-0105537.                                              |
| XX     |                                                                         |
| XX     | (MINU ) UNIV MINNESOTA.                                                 |

XX Sherman DH, Liu H, Xue Y, Zhao L;  
 XX WPI: 2000-160679/14.  
 XX P-PSDB; AAY77181.  
 XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.  
 XX synthesis of methymycin and pikromycin -  
 XX Claim 3; Page 353-354; 438pp; English.  
 XX The invention relates to an isolated and purified nucleic acid segment  
 XX comprising a desosamine biosynthetic gene cluster, a fragment or its  
 XX biologically active variant, where the nucleic acid sequence is not  
 XX derived from the eryC gene cluster of *Saccharopolyspora erythraea* or  
 XX Streptomyces antibiotics. The invention also relates to a macrolide  
 XX biosynthetic gene cluster, or fragments thereof. The macrolide  
 XX biosynthetic gene cluster encodes proteins which synthesize methymycin,  
 XX pikromycin, neomethymycin, narbomycin or a combination of these  
 XX compounds. Recombinant or augmented cells comprising the desosamine  
 XX and/or macrolide biosynthetic gene clusters are useful for the  
 XX production of biologically active macrolides. The macrolide biosynthetic  
 XX proteins are useful for synthesis of methymycin, pikromycin,  
 XX neomethymycin and narbomycin. The alternative termination of polyketide  
 XX synthesis may be useful to prepare novel antibiotics and  
 XX polyhydroxyalkanoate (PHA) monomers. The compounds produced by the  
 XX recombinant host cells are useful as biopolymers, e.g., in packaging or  
 XX biomedical applications, to engineer PHA monomer synthases or to prepare  
 XX biologically active agents, such as chemotherapeutics,  
 XX immunosuppressants, agents to treat asthma, chronic obstructive pulmonary  
 XX disease as well as other diseases involving respiratory inflammation,  
 XX cholesterol-lowering agents or macrolide-based antibiotics which are  
 XX active against a variety of organisms, e.g., bacteria, including  
 XX multi-drug resistant pneumococci and other respiratory pathogens, as well  
 XX as viral parasitic pathogens, or as crop protection agents (e.g.,  
 XX fungicides or insecticides) via expression of polyketides in plants.  
 XX Sequences AA87286-287294 represent desosamine biosynthetic genes from  
 XX Streptomyces venezuelae ATCC 15439, which encode proteins  
 XX AAY77181-Y77189.  
 XX Sequence 1248 BP: 177 A; 546 C; 377 G; 148 T; 0 other;  
 XX  
 XX Query Match 71.5%; Score 18.6; DB 21; Length 1248;  
 XX Best Local Similarity 84.0%; Pred. No. 29;  
 XX Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 XX  
 XX QY 1 cccgtctaggcgttcggtgtccggc 25  
 XX Db 1068 CGCGCGAGCGGCTTCGGTGTGCGGC 1044  
 XX  
 XX RESULT 2  
 XX AAZ87284  
 XX ID AAZ87284 standard; DNA; 12441 BP.  
 XX AC AAZ87284;  
 XX XX 05-JUN-2000 (first entry)  
 XX DE S. venezuelae desosamine biosynthetic gene cluster pikB, SEQ ID NO:3.  
 XX XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;  
 XX KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;  
 XX KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,  
 XX KW chronic obstructive pulmonary disease; respiratory inflammation;  
 XX KW hypercholesterolaemia; crop protection agent; ds.  
 XX XX Streptomyces venezuelae ATCC15439.  
 XX OS W0200000620-A2.  
 XX XX W0200000620-A2.  
 XX XX 06-JAN-2000.  
 XX PD

XX 25-JUN-1999; 99WO-US14398.  
 XX 26-JUN-1998; 98US-0105537.  
 XX (MINU ) UNIV MINNESOTA.  
 XX Sherman DH, Liu H, Xue Y, Zhao L;  
 XX WPI: 2000-160679/14.  
 XX P-PSDB; AAY77179.  
 XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.  
 XX synthesis of methymycin and pikromycin -  
 XX Claim 2; Page 281-287; 438pp; English.  
 XX The invention relates to an isolated and purified nucleic acid segment  
 XX comprising a desosamine biosynthetic gene cluster, a fragment or its  
 XX biologically active variant, where the nucleic acid sequence is not  
 XX derived from the eryC gene cluster of *Saccharopolyspora erythraea* or  
 XX Streptomyces antibiotics. The invention also relates to a macrolide  
 XX biosynthetic gene cluster, or fragments thereof. The macrolide  
 XX biosynthetic gene cluster encodes proteins which synthesize methymycin,  
 XX pikromycin, neomethymycin, narbomycin or a combination of these  
 XX compounds. Recombinant or augmented cells comprising the desosamine  
 XX and/or macrolide biosynthetic gene clusters are useful for the production  
 XX of biologically active macrolides. The macrolide biosynthetic proteins  
 XX are useful for synthesis of methymycin, pikromycin, neomethymycin and  
 XX narbomycin. The alternative termination of polyketide synthesis may be  
 XX useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)  
 XX monomers. The compounds produced by the recombinant host cells are useful  
 XX as biopolymers, e.g., in packaging or biomedical applications, to  
 XX engineer PHA monomer synthases or to prepare biologically active agents,  
 XX such as chemotherapeutics, immunosuppressants, agents to treat asthma,  
 XX chronic obstructive pulmonary disease as well as other diseases involving  
 XX respiratory inflammation, cholesterol-lowering agents or macrolide-based  
 XX antibiotics which are active against a variety of organisms, e.g.,  
 XX bacteria, including multi-drug resistant pneumococci and other  
 XX respiratory pathogens, as well as viral parasitic pathogens, or as crop  
 XX protection agents (e.g., fungicides or insecticides) via expression of  
 XX polyketides in plants. The present sequence represents the desosamine  
 XX biosynthetic gene cluster from Streptomyces venezuelae ATCC 15439.  
 XX Sequence 12441 BP: 1704 A; 4294 C; 4686 G; 1757 T; 0 other;  
 XX  
 XX Query Match 71.5%; Score 18.6; DB 21; Length 12441;  
 XX Best Local Similarity 84.0%; Pred. No. 31;  
 XX Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 XX  
 XX QY 1 cccgtctaggcgttcggtgtccggc 25  
 XX Db 11129 cgcgcgcgcgcgttcggtgtccggc 11153  
 XX  
 XX RESULT 3  
 XX AAZ87319  
 XX ID AAZ87319 standard; DNA; 13613 BP.  
 XX AC AAZ87319;  
 XX XX 05-JUN-2000 (first entry)  
 XX DE S. venezuelae desosamine biosynthetic gene cluster pikB.  
 XX KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;  
 XX KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;  
 XX KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,  
 XX KW chronic obstructive pulmonary disease; respiratory inflammation;  
 XX KW hypercholesterolaemia; crop protection agent; ds.  
 XX XX Streptomyces venezuelae ATCC15439.  
 XX OS

```
XX Key Location/Qualifiers
FH CDS 1..809
FT /tag= a
FT /partial
FT /product= "PikB gene cluster protein #1 (AAV77204)"
FT /note= "No initiation codon given in the specification"
FT CDS 806..2014
FT /tag= b
FT /product= "PikB gene cluster protein #2 (AAV80998)"
FT 2162..6741
FT /tag= c
FT /product= "PikB gene cluster protein #3 (AAV77205)"
FT 6834..7402
FT /tag= d
FT /product= "PikB gene cluster protein #4 (AAV77206)"
FT /transl_except= (pos:6837..6841, aa:Gln)
FT CDS 7492..8205
FT /tag= e
FT /product= "PikB gene cluster protein #5 (AAV77207)"
FT complement (7942..8205)
FT /partial
FT /tag= f
FT /product= "PikB gene cluster protein #6 (AAV77208)"
FT /note= "No termination codon given in the specification"
FT /transl_except= (pos:8270..8272, aa:Val)
FT /transl_except= (pos:8273..8275, aa:Thr)
FT /transl_except= (pos:8276..8278, aa:Gly)
FT CDS complement (10126..11139)
FT /tag= g
FT /product= "PikB gene cluster protein #7 (AAV80999)"
FT complement (11271..12149)
FT /tag= h
FT /product= "PikB gene cluster protein #8 (AAV77209)"
FT complement (12342..13799)
FT /tag= i
FT /product= "PikB gene cluster protein #9 (AAV77210)"
FT complement (13706..15043)
FT /tag= j
FT /product= "PikB gene cluster protein #10 (AAV77211)"
FT 13404..15574
FT /tag= k
FT /product= "PikB gene cluster protein #11 (AAV77212)"
XX WO200000620-A2.
XX
XX 06-JAN-2000.
XX
XX 25-JUN-1999; 99WO-US14398.
XX
XX 26-JUN-1998; 98US-0105537.
XX (MINU ) UNIV MINNESOTA.
XX
XX Sherman DH, Liu H, Xue Y, Zhao L;
XX
XX WPI: 2000-160679/14.
XX
XX P-PSDB: AAV77204, AAV77205, AAV77207, AAV77207, AAV77208.
XX AAV77209, AAV77210, AAV77211, AAV77212, AAV80998, AAV80999.
XX
XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
XX synthesis of methymycin and pikromycin -
XX
XX Disclosure; Figure 32; 438pp; English.
XX
XX The invention relates to an isolated and purified nucleic acid segment
XX comprising a desosamine biosynthetic gene cluster, a fragment or its
XX biologically active variant, where the nucleic acid sequence is not
XX derived from the eryC gene cluster of Saccharopolyspora erythraea or
XX Streptomyces antibioticus. The invention also relates to a macrolide
XX biosynthetic gene cluster, or fragments thereof. The macrolide
XX biosynthetic gene cluster encodes proteins which synthesise methymycin,
XX pikromycin, neomethymycin, narbomycin or a combination of these
```

```
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the production
CC of biologically active macrolides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC narbomycin. The alternative termination of polyketide synthase may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to
CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents the desosamine
CC biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC
CC 15439, as given in figure 32.
XX
XX Sequence 13613 BP; 1858 A; 4733 C; 5092 G; 1930 T; 0 other;
SQ
Query Match 71.5%; Score 18.6; DB 21; Length 13613;
Best Local Similarity 84.0%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 cccgtctaggcgttcggtgcgcgc 25
Db 11816 cgcgcgcgcgcgttcggtgcgcgc 11840
RESULT 4
AAF30757/C
ID AAF30757 standard; DNA; 47981 BP.
XX
XX AAF30757;
XX
XX 21-JUN-2001 (first entry)
XX
XX Micromonospora megalomicea megalomicin biosynthetic gene cluster.
KW Megalomicin; meg gene; polyketide synthase; antibiotic;
KW motilide; antiparasitic; ds.
XX
XX Micromonospora megalomicea subsp. nigra.
XX
XX Key Location/Qualifiers
FH CDS complement (1..144)
FT /tag= a
FT /partial
FT /gene= "megBVI(megT)"
FT /product= "TDP-4-keto-6-deoxyglucose-
FT 2,3-dehydratase"
FT 928..2061
FT /note= "encodes AAB82201"
FT CDS
FT 928..2061
FT /tag= b
FT /gene= "megDVI"
FT /product= "TDP-4-keto-6-deoxyhexose 3,4-isomerase"
FT 2072..3382
FT /note= "encodes AAB82202"
FT CDS
FT 2072..3382
FT /tag= c
FT /gene= "megDI"
FT /product= "TDP-megosamine glycosyltransferase"
FT /note= "eryCIII homologue; encodes AAB802203"
FT 3462..4634
FT /tag= d
FT /gene= "megY"
FT /product= "mycarose O-acyltransferase"
FT 4651..5775
FT /note= "encodes AAB82204"
FT CDS
FT 4651..5775
FT /tag= e
FT /gene= "megDII"
```

```
FT FT /product= "TDP-3-keto-6-deoxyhexose
FT FT 3-aminotransaminase"
FT FT /note= "eryCI, DnrJ homologue, encodes AAB82205"
FT FT 5822..6595
FT FT /*tag= f
FT FT /gene= "megDIII"
FT FT /product= "daunosaminyl-N,N-dimethyltransferase"
FT FT /note= "eryCVI homologue; encodes AAB82206"
FT FT 6592..7197
FT FT /*tag= g
FT FT /gene= "megDIV"
FT FT /product= "TDP-4-keto-6-deoxyhexose 3,5-epimerase"
FT FT /note= "eryVIII, dnmV homologue; encodes AAB82207"
FT FT 7220..8206
FT FT /*tag= h
FT FT /gene= "megDV"
FT FT /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
FT FT /note= "eryVIV, dnmV homologue; encodes AAB82208"
FT FT complement (8228..9220)
FT FT /*tag= i
FT FT /gene= "megBII-1(megDVII)"
FT FT /product= "TDP-4-keto-6-deoxyhexose 2,3-reductase"
FT FT /note= "encodes AAB82209"
FT FT complement (9226..10479)
FT FT /*tag= j
FT FT /gene= "megBV"
FT FT /product= "TDP-mycarose glycosyltransferase"
FT FT /note= "encodes AAB82210"
FT FT complement (10483..11424)
FT FT /*tag= k
FT FT /gene= "megBIV"
FT FT /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
FT FT /note= "encodes AAB82211"
FT FT 12181..22821
FT FT /gene= "megA"
FT FT /*tag= l
FT FT /product= "megalomicin 6-deoxyerythronolide B
FT FT synthase 1"
FT FT /note= "polyketide synthase; encodes AAB82212"
FT FT 12505..13470
FT FT /*tag= m
FT FT /gene= "megA"
FT FT /function= "AT-L"
FT FT 13576..13791
FT FT /*tag= n
FT FT /gene= "megA"
FT FT /function= "ACP-L"
FT FT 13849..15126
FT FT /*tag= o
FT FT /gene= "megA"
FT FT /function= "KS1"
FT FT 15427..16476
FT FT /*tag= p
FT FT /gene= "megA"
FT FT /function= "Atl"
FT FT 17155..17694
FT FT /*tag= q
FT FT /gene= "megA"
FT FT /function= "KR1"
FT FT 17947..18207
FT FT /*tag= r
FT FT /gene= "megA"
FT FT /function= "ACP1"
FT FT 18268..19548
FT FT /*tag= s
FT FT /gene= "megA"
FT FT /function= "KS2"
FT FT 19876..20910
FT FT /*tag= t
FT FT /gene= "megA"
FT FT /function= "AT2"
FT FT 21517..22053
FT FT /*tag= u
FT FT /gene= "megA"
FT FT /function= "KR2"
FT FT 22318..22575
FT FT /*tag= v
FT FT /gene= "megA"
FT FT /function= "ACP2"
FT FT 22867..33555
FT FT /*tag= w
FT FT /gene= "megAII"
FT FT /product= "megalomicin 6-deoxyerythronolide B
FT FT synthase 2"
FT FT /note= "polyketide synthase, encodes AAB82213"
FT FT 22957..24237
FT FT /*tag= x
FT FT /gene= "megAII"
FT FT /function= "KS3"
FT FT 24544..25581
FT FT /*tag= y
FT FT /gene= "megAII"
FT FT /function= "AT3"
FT FT 26230..26733
FT FT /*tag= z
FT FT /gene= "megAII"
FT FT /function= "KR3 (inactive)"
FT FT 26998..27258
FT FT /*tag= aa
FT FT /gene= "megAII"
FT FT /function= "ACP3"
FT FT 27393..28590
FT FT /*tag= ab
FT FT /gene= "megAII"
FT FT /function= "KS4"
FT FT 28897..29931
FT FT /*tag= ac
FT FT /gene= "megAII"
FT FT /function= "AT4"
FT FT 29953..30477
FT FT /*tag= ad
FT FT /gene= "megAII"
FT FT /function= "DH4"
FT FT 31396..32244
FT FT /*tag= ae
FT FT /gene= "megAII"
FT FT /function= "ER4"
FT FT 32257..32799
FT FT /*tag= af
FT FT /gene= "megAII"
FT FT /function= "KR4"
FT FT 33052..33312
FT FT /*tag= ag
FT FT /gene= "megAII"
FT FT /function= "ACP4"
FT FT 33666..43271
FT FT /*tag= ah
FT FT /gene= "megAII"
FT FT /product= "megalomicin 6-deoxyerythronolide B
FT FT synthase 3"
FT FT /note= "polyketide synthase; encodes AAB82214"
FT FT 22957..24237
FT FT /*tag= ai
FT FT /gene= "megAII"
FT FT /function= "KS5"
FT FT 24544..25581
FT FT /*tag= aj
FT FT /gene= "megAII"
FT FT /function= "AT5"
FT FT 26230..26733
FT FT /*tag= ak
FT FT /gene= "megAII"
FT FT /function= "KR5"
FT FT 26998..27258
FT FT /*tag= al
FT FT /gene= "megAII"
```

FT misc\_feature /function= "ACP5"  
FT 27393..28590  
FT \*tag= am  
FT /gene= "megAIII"  
FT /function= "K36"  
FT 28897..29931  
FT \*tag= an  
FT /gene= "megAIII"  
FT /function= "AT6"  
FT misc\_feature 29953..30477  
FT \*tag= ao  
FT /gene= "megAIII"  
FT /function= "KR6"  
FT misc\_feature 31396..32244  
FT \*tag= ap

Query Match 67.7%; Score 17.6; DB 22; Length 47981;  
Best Local Similarity 83.3%; Pred. NO. 92;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cgcctaggcgttcggtccggc 25  
|||||  
DB 5681 CCGGCCAGCGTTCGGTACCGGC 5658

RESULT 5  
AAH41415/C  
ID AAH41415 standard; cDNA; 1680 BP.

XX AC AAH41415;

XX XX 23-AUG-2001 (first entry)

XX DE Human pentraxin I encoding cDNA SEQ ID NO:1.

XX KW Human; pentraxin I; pentraxin receptor; neuronal disorder; inhibitor;  
XX KW detection; neuroprotective; nontropic; cerebroprotective; cytostatic;  
XX KW gene therapy; acute head trauma; multiple sclerosis; spinal cord injury;  
XX KW Alzheimer's disease; brain tumour; stroke; neuroprotective; ss.

XX OS Homo sapiens.

XX PN EF1101820-A1.

XX XX 23-MAY-2001.

XX XX 15-NOV-1999; 99EP-0122666.

XX XX 15-NOV-1999; 99EP-0122666.

XX PA (MERE ) MERCK PATENT GMBH.

XX PI Messegue Peypoch R, Rosell Vives E, Martinez Escola JM;

XX PI Rodas Gubern B, Adan Plana J, Puig CN, Carceller RA;

XX PI Masa Alvarez M, Piuats X, Den Daas I, Trullas OR;

XX PI DeGregorio-Rocasolano BN;

XX DR WPI; 2001-357928/38.

XX PT New nucleic acid encoding the human Pentraxin receptor useful for  
PT preventing, treating, ameliorating or detecting a disease associated  
PT with a neuronal disorder such as a stroke, multiple sclerosis, or  
PT Alzheimer's disease

XX PS Example 2; Fig 17; 53pp; English.

XX CC The present invention describes a nucleic acid (I) encoding the human  
CC pentraxin receptor (hPr) or a protein exhibiting biological properties  
CC of hPr and having a sequence selected from: (a) a fully defined 1853 bp  
CC sequence given in AAH41415; (b) a sequence which hybridises with (a);  
CC (c) a degenerate variant which deviates from (a) or (b); and (d) a  
CC fragment, derivative or allelic variation of (a), (b) or (c). (I) has  
CC neuroprotective, nontropic, cerebroprotective and cytostatic activity,

CC and can be used in gene therapy. Pentraxin I, hPr, or hPr-encoding  
CC sequences are useful in the preparation of pharmaceutical compositions  
CC for preventing, treating or ameliorating a disease associated with a  
CC neuronal disorder such as stroke, acute head trauma, multiple sclerosis,  
CC spinal cord injury, Alzheimer's disease or brain tumour, or for  
CC providing a neuroprotective effect. The present sequence represents the  
CC human cDNA coding sequence of pentraxin I, which is used in an example  
CC from the present invention.

XX SQ Sequence 1680 BP; 328 A; 571 C; 526 G; 255 T; 0 other;

Query Match 66.2%; Score 17.2; DB 22; Length 1680;  
Best Local Similarity 86.4%; Pred. NO. 1.3e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 cgcctaggcgttcggtccggc 24  
|||||  
DB 519 CGTCTCGCGCGCGTCTCCGG 498

RESULT 6  
ABA46067/C

ID ABA46067 standard; DNA; 381 BP.

XX AC ABA46067;

XX XX 01-FEB-2002 (first entry)

XX DE Human breast cell single exon nucleic acid probe #4762.

XX KW Human; microarray; single exon probe; gene expression; breast;  
XX KW disease; cancer; ss.

XX OS Homo sapiens.

XX PN WO200157271-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00662.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX XX WPI; 2001-496933/54.

XX DR New spatially-addressable set of single exon nucleic acid probes,  
XX PT useful for measuring gene expression in sample derived from human  
XX PT breast, comprises number of single exon nucleic acid probes

XX PS Claim 1; SEQ ID NO 4762; 327pp + sequence listing; English.

XX CC The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human breast and BT 474 cells. The method involves contacting  
CC the probes with a collection of detectably labelled nucleic acids  
CC derived from mRNA of human breast, and then measuring the label  
CC bound to each probe of the microarray. The probes are useful for  
CC verifying the expression of regions of genomic DNA predicted to  
CC encode proteins. They are useful for gene discovery, and for  
CC determining predisposition and/or prognosing breast disease. Gene  
CC expression analysis is useful for assessing the toxicity of chemical  
CC agents on cells. The microarray of this invention presents a far greater

CC diversity of probes for measuring gene expression, with far less bias  
CC than expressed sequence tag microarrays. The method is suitable for  
CC rapid production of functional information from genomic sequence. The  
CC present sequence is a single exon nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 381 BP; 92 A; 130 C; 95 G; 64 T; 0 other;

Query Match 65.4%; Score 17; DB 22; Length 381;

Best Local Similarity 80.0%; Pred. No. 1.5e+02;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ccgtctaggcgttcggtccggcc 26

Db 241 CGGGCGAGCGTTCAGAGTCGGCC 217

RESULT 7

ABA56605/C

ID ABA56605 standard; DNA; 381 BP.

XX AC ABA56605;

XX 01-FEB-2002 (first entry)

XX Human foetal liver single exon nucleic acid probe #4910.

XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human fetal liver -

XX Claim 1; SEQ ID NO 4910; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for  
XX measuring human gene expression in a sample derived from human foetal  
XX liver. The single exon nucleic acid probes may be used for predicting,  
XX measuring and displaying gene expression in samples derived from human  
XX fetal liver. The present sequence is a single exon nucleic acid  
XX probe of the invention.

XX Note: The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 381 BP; 92 A; 130 C; 95 G; 64 T; 0 other;

Query Match

65.4%; Score 17; DB 22; Length 381;

Best Local Similarity 80.0%; Pred. No. 1.5e+02;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ccgtctaggcgttcggtccggcc 26

Db 241 CGGGCGAGCGTTCAGAGTCGGCC 217

RESULT 8

ABA26220/C

ID ABA26220 standard; DNA; 381 BP.

XX AC ABA26220;

XX 23-JAN-2002 (first entry)

XX Probe #4686 for gene expression analysis in human heart cell sample.

XX Human; gene expression; heart; microarray; vascular system; probe;

XX cardiovascular disease; hypertension; cardiac arrhythmia;

XX congenital heart disease; ss.

XX Homo sapiens.

XX WO200157274-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00666.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488899/53.

XX Single exon nucleic acid probes for analyzing gene expression in human  
XX hearts -

XX Claim 1; SEQ ID No 4686; 530pp; English.

XX The present invention relates to single exon nucleic acid probes for  
XX measuring human gene expression in a sample derived from human heart. The  
XX present sequence is one such probe. The probes may be used for  
XX predicting, measuring and displaying gene expression in samples derived  
XX from the human heart via microarrays. By measuring gene expression, the  
XX probes are useful for predicting, diagnosing, grading, staging,  
XX monitoring and prognosing diseases of the human heart and vascular system  
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
XX congenital heart disease.

XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 381 BP; 92 A; 130 C; 95 G; 64 T; 0 other;

Query Match

65.4%; Score 17; DB 22; Length 381;

Best Local Similarity 80.0%; Pred. No. 1.5e+02;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ccgtctaggcgttcggtccggcc 26

Db 241 CGGGCGAGCGTTCAGAGTCGGCC 217









CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
 CC the proteins are collectively known as colon cancer antigens. The colon  
 CC cancer antigens have cytostatic activity and can be used in gene  
 CC therapy and vaccine production. N and P may be used in the prevention,  
 CC diagnosis and treatment of diseases associated with inappropriate P  
 CC expression. For example, N and P may be used to treat disorders  
 CC associated with decreased expression by rectifying mutations or deletions  
 CC in a patient's genome that affect the activity of P by expressing  
 CC inactive proteins or to supplement the patient's own production of P.  
 CC Additionally, N may be used to produce the colon cancer-associated ps,  
 CC by inserting the nucleic acids into a host cell and culturing the cell  
 CC to express the proteins. N and P can be used in the prevention, diagnosis  
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
 CC and AAB77789 represent sequences used in the exemplification of the  
 CC present invention.  
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
 CC missing at time of publication, meaning no sequences are present for  
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
 XX  
 SQ Sequence 2443 BP; 612 A; 571 C; 599 G; 659 T; 2 other;

Query Match 65.4%; Score 17; DB 22; Length 2443;  
 Best Local Similarity 80.0%; Pred. No. 1.6e+02;  
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Qy 2 ccgtctaggcgttcggtgcggcc 26  
 ||| |||| |||| ||||  
 Db 130 ccgmctagccgtgcggtgccagcc 154

Search completed: July 31, 2002, 14:11:22  
 Job time: 17379 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:01:44 ; Search time 6034.22 Seconds  
(without alignments)  
90.167 Million cell updates/sec

Title: US-09-899-718A-5  
Perfect score: 26  
Sequence: 1 cccgtcattagcgttcgtgtccgcc 26

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797556 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_hgt.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pi.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_hgt\_hum.\*
- 31: em\_hgt\_inv.\*
- 32: em\_hgt\_other.\*
- 33: em\_hgtg\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No.                                                                                                                                                                                                | Score | Match | Length | DB | ID | Description |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|-------|--------|----|----|-------------|
| <p>Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.</p> |       |       |        |    |    |             |

|    |      |       |        |   |           |
|----|------|-------|--------|---|-----------|
| 1  | 26   | 100.0 | 26     | 6 | AX349067  |
| 2  | 26   | 100.0 | 3785   | 6 | AX349063  |
| 3  | 19.8 | 76.2  | 30150  | 1 | SC6D10    |
| 4  | 18.6 | 71.5  | 12441  | 1 | AF079762  |
| 5  | 18.6 | 71.5  | 37969  | 3 | AC005928  |
| 6  | 18.6 | 71.5  | 41944  | 3 | AC005929  |
| 7  | 18.2 | 70.0  | 1432   | 4 | CF096127  |
| 8  | 18.2 | 70.0  | 1466   | 4 | AF045773  |
| 9  | 18.2 | 70.0  | 5908   | 1 | AF147704  |
| 10 | 18.2 | 70.0  | 22396  | 1 | SCF43     |
| 11 | 18.2 | 70.0  | 173434 | 2 | AC098934  |
| 12 | 18   | 69.2  | 3149   | 1 | SLMELOPER |
| 13 | 18   | 69.2  | 6927   | 1 | AF319543  |
| 14 | 18   | 69.2  | 20633  | 3 | AC109873  |
| 15 | 18   | 69.2  | 40064  | 3 | LMFL6586  |
| 16 | 18   | 69.2  | 45653  | 2 | AC105239  |
| 17 | 18   | 69.2  | 61843  | 8 | NC15E6    |
| 18 | 18   | 69.2  | 79133  | 8 | NC15E11   |
| 19 | 18   | 69.2  | 122393 | 2 | OSJN00101 |
| 20 | 17.8 | 68.5  | 11355  | 1 | AE008357  |
| 21 | 17.8 | 68.5  | 12045  | 1 | AE009255  |
| 22 | 17.6 | 67.7  | 41097  | 1 | AF016585  |
| 23 | 17.6 | 67.7  | 47981  | 1 | AF263245  |
| 24 | 17.6 | 67.7  | 47981  | 6 | AX112026  |
| 25 | 17.6 | 67.7  | 165812 | 2 | AP001141  |
| 26 | 17.6 | 67.7  | 166063 | 9 | AP002515  |
| 27 | 17.6 | 67.7  | 168213 | 2 | AC103175  |
| 28 | 17.6 | 67.7  | 168308 | 2 | AC084775  |
| 29 | 17.6 | 67.7  | 175464 | 2 | AP002887  |
| 30 | 17.6 | 67.7  | 185800 | 2 | AC079366  |
| 31 | 17.4 | 66.9  | 115948 | 2 | AC094962  |
| 32 | 17.4 | 66.9  | 146742 | 2 | AL139254  |
| 33 | 17.2 | 66.2  | 1668   | 9 | HSU48252  |
| 34 | 17.2 | 66.2  | 1680   | 6 | AX137782  |
| 35 | 17.2 | 66.2  | 1680   | 6 | AX148704  |
| 36 | 17.2 | 66.2  | 2059   | 9 | BC018289  |
| 37 | 17.2 | 66.2  | 2070   | 9 | HSM600858 |
| 38 | 17.2 | 66.2  | 2286   | 9 | AF006013  |
| 39 | 17.2 | 66.2  | 2396   | 9 | HSU49262  |
| 40 | 17.2 | 66.2  | 3403   | 9 | HSU75651  |
| 41 | 17.2 | 66.2  | 5071   | 9 | HSU61849  |
| 42 | 17.2 | 66.2  | 5146   | 9 | D86963    |
| 43 | 17.2 | 66.2  | 5966   | 1 | SAMT18862 |
| 44 | 17.2 | 66.2  | 189050 | 1 | AL646077  |
| 45 | 17.2 | 66.2  | 193068 | 8 | AP003455  |

## ALIGNMENTS

RESULT 1  
AX349067  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
Source

AX349067  
Sequence 5 from Patent WO0202785.  
AX349067  
AX349067.1 GI:18615102  
synthetic construct.  
synthetic construct.  
artificial sequence.  
1 (sites)  
Spruncker, S., Kluth, A., Becker, D., Luetticke, S. and Loerz, H.  
Promoters of gene expression in plant caryopses  
Patent: WO 0202785-A 5 10-JAN-2002;  
Aventis CropScience GmbH (DE)  
Location/Qualifiers  
1..26  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="Oligonucleotide"  
1 a 10 c 9 g 6 t

BASE COUNT  
ORIGIN

Query Match 100.0%; Score 26; DB 6; Length 26;  
Best Local Similarity 100.0%; Pred. No. 0.53;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cccgttagcggttcggtccggcc 26  
|||||  
DB 1 cccgttagcggttcggtccggcc 26

RESULT 2  
AX349063  
LOCUS AX349063 3785 bp DNA linear PAT 06-FEB-2002  
DEFINITION Sequence 1 from Patent WO0202785.  
ACCESSION AX349063  
VERSION AX349063.1 GI:18615098  
KEYWORDS bread wheat.  
SOURCE Triticum aestivum  
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae; Triticeae; Triticum.  
REFERENCE 1 (sites)  
AUTHORS Sprunck,S., Kluth,A., Becker,D., Luetticke,S. and Loerz,H.  
TITLE Promoters of gene expression in plant caryopses  
JOURNAL Patent: WO 0202785-A 1 10-JAN-2002;  
Aventis CropScience GmbH (DE)  
LOCATION/Qualifiers  
FEATURES  
source  
1. .3785  
/organism="Triticum aestivum"  
/db\_xref="taxon:4565"  
BASE COUNT 945 a 980 c 899 g 961 t  
ORIGIN

Query Match 100.0%; Score 26; DB 6; Length 3785;  
Best Local Similarity 100.0%; Pred. No. 0.44;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cccgttagcggttcggtccggcc 26  
|||||  
DB 2107 CCGTCTAGCGGTCGGTGTCCGGCC 2132

RESULT 3  
SC6D10  
LOCUS SC6D10 30150 bp DNA linear BCT 01-FEB-2000  
DEFINITION Streptomyces coelicolor cosmid 6D10.  
ACCESSION AL138538  
VERSION AL138538.1 GI:6855381  
KEYWORDS ABC transporter ATP-binding protein; globin; kinase; lipoprotein; membrane protein; O-methyltransferase; protein kinase; secreted protein; secreted sugar-binding protein; sugar hydrolase; sugar transport inner membrane protein; sugar transport system permease protein; sugar-binding protein.  
SOURCE Streptomyces coelicolor A3(2).  
ORGANISM Streptomyces coelicolor A3(2)  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 30150)  
AUTHORS Redenbach,M., Kieser,H.M., Denapate,D., Eichner,A., Cullum,J., Kinashi,H. and Hopwood,D.A.  
TITLE A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)  
MEDLINE 97000351  
REFERENCE 2 (bases 1 to 30150)  
AUTHORS Seeger,K.J. and Harris,D.  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 30150)  
AUTHORS Cerdano,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.  
TITLE Direct Submission  
JOURNAL Submitted (01-FEB-2000) Streptomyces coelicolor sequencing project,

COMMENT

Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

Notes:  
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics  
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
(URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>)  
CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).  
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.  
The length in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.  
The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.  
Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.  
Cosmid 6D10 overlaps with cosmid C61 on the AseI-C genomic restriction fragment.

FEATURES  
source

1. .30150  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
/clone="cosmid 6D10"  
1. .823  
/gene="SC6D10.01"  
1. .823  
/gene="SC6D10.01"  
/note="Protein sequence is in conflict with the conceptual translation: SC6D10.01, possible sugar-binding protein (partial) len: > 273 aa; similar to TR:CAB49101 (EMBL:AU248283) Pyrococcus abyssi maltose-binding periplasmic protein precursor MalE, 453 aa; fasta scores: opt: 246 z-score: 288.7 E(): 1.1e-08; 24.0% identity in 271 aa overlap and to SW:MALX\_STRPN (EMBL:L08611) Streptococcus pneumoniae maltose/maltodextrin-binding protein precursor MalX, 423 aa; fasta scores: opt: 184 z-score: 217.6 E(): 0.0001; 22.9% identity in 279 aa overlap  
/codon\_start=2  
/transl\_table=11  
/product="putative sugar-binding protein"  
/protein\_id="CAB71196.1"  
/db\_xref="GI:6855382"  
/translation="MKAPPPTWSEFAQVAKLTGKDGYPOLGFMFNVHYVTVVDH YMSQMDHAYFDEGRSAEDPAFEMFTYQKLLVDLGGEEKLEKRYNTFGDEMGAK HPFQTQGVAMQLDGEWRLGMADGAVDFEIGTAPMPVADDAEYTGKGLSGTVMGIA PQSKQNAWELVKYMTTDTGAVVAFANIRNVPFPALSKPLDKTDPDKFTFLDIA QHRESNPPASVNGATYQLTLQDLGYQYESGKVKDKLKGLEKAAQIDRDIEQAK"  
1. .100  
/gene="SC6D10.01"  
/note="nominal overlap with S. coelicolor cosmid St854A"  
830. .1774  
/gene="SC6D10.02"

gene  
CDS  
misc\_feature  
gene



```

/product="putative tautomerase"
/protein_id="AAC68677.1"
/db_xref="GI:3789893"
/translation="WTDDLGTALTOPPLGRVTRAVADRELCTHLETRGHIHHAANG
DPYAVYLRQADDPYPAYERVARGALSFSTGSGVTVADHAAASILCSDFGVSGAD
GVFVQVLSGEGCPLEREQVLPFAAGDPEGGQVAVVAGHRETLGLADPDSASTA
FELLGGVPVAVTAAAVAVGVADRRADFADLLERPLSLLSLLAQPSLRTVRAAG
ALLETALLADSDSPGALSALGYTAQVLTGNVALLAHLPQWRELCDPGLAAA
AVETLRDYPVQIDARVVRGTELAGRLRFPAGAVVVLTAATGRDPEVFDPERFDL
APEDAHAHLHPAGPYGPVASVLRLEVALRTLAGRFPGLRQAGDVLRRPVPGR
GPLSVPVSS"
gene
1475..2755
/gene="desVII"
CDS
1475..2755
/gene="desVII"
/codon_start=1
/transl_table=11
/product="glycosyl transferase"
/protein_id="AAC68677.1"
/db_xref="GI:3789894"
/translation="MRVLTGSAFHHTHYGYGLVPLAWALLAAGHEVVRVASQPALTDIT
GSLAAPVGTGDLHLEYVRMAGPRNHPAIAFDEARPEPLDWDHAGLGEAILAPY
FHLANNDSWDDLVDFARSQPDVLWPEPTTYAGAVAAQVTGAHARVLMGPDVMS
ARKKVALRDQPPHREDPTAEWLITWLDKYGASFEEELLTGQFTIDPTPSLRIDT
GLPTVMRYPTNGTSVVDMLSEPPARPVCLTUGVSAREVLGQDGVSGQDILEALA
DLDIELVATLDASQAEIRNPKHRTFTDFPMHALLPSCSAIHHGGAGTYATAVIN
AVPQMLAELMDAPYKARAVABQAGFELPPAELTPQAVRDVVRILDDPSVATAAHR
LREETFGDTPAGIVPELERLAAQHRPPADARH"
gene
2845..3558
/gene="desVI"
CDS
2845..3558
/gene="desVI"
/codon_start=1
/transl_table=11
/product="N,N-dimethyltransferase"
/protein_id="AAC68678.1"
/db_xref="GI:3789895"
/translation="MYEVDHADVDYDLFVLRGKDYAAEASDIADLRVTRTPEASLLD
VACGTGLEHTKEFTAGLESDMLTHARKRLPATLHQGDMRDFRLGRKFSAV
VSMFSVGLYKTEELGAASFAHLEPGGVVVEPWPFETFDAGWVSDVVRDQ
RIVARVSVRGNATREVEHFTVADPGKGVHRFSDVHLITLTHQAEYEAATAAGLR
VEYLEGSGRGLFVGVA"
gene
3625..6054
/gene="desR"
CDS
3625..6054
/gene="desR"
/translation="putative extracellular protein involved in a
resistance mechanism"
/codon_start=1
/transl_table=11
/product="beta-glucosidase"
/protein_id="AAC68679.1"
/db_xref="GI:3789896"
/translation="WTGKTRIPRVRRGRTTPRAFTLAVVGTLLAGTTVAALAPGAADT
ANQVTSRAELVAOMTLDKISFVHWALDDPRNVGYLPVGPRLGIPELRAADPCNG
IRLGGTALPAPALVASTTDFDTWADSTGKVMGRDGRALNODMVLGPMNNRVPFG
GRNYETSDPLVSRATAVAQIKGAGLMTAKHFAANQNNRFSVNAVNDQTL
RIETFPFAFSAKGAASFCMAYNGLNGKPCSGNDELNNVLTQWQGVWMSDLA
TPGDAITKGLQDMGVLPDVPKGPSPPAKFFGEALKTAVLNGTVPEAAVTRSAE
RIVGMEKFKGLLATPAPRPDRKAGAAQVSRKVAENGAVILLRNEGALPLAGDAGKS
LIAVGTAVDVKVGTGLGSAHVVPDSAAAPLDTIKARAGATVTVETGEFTGQIPA
GNLSPAFNGHOLEFGKAGALXDGLTVPADGEYRIARVATGGYATVQLGSHTEAGQ
VYKVSPPLLKLTIKGTHLITSGFAMSATPLSLELGWVTPPAADATIAKAVESAKAR
TAVFAYDDGTGVRPNLSLPGTQDKLISAVADANPNTIYVLTNGSSVLPWLSKTR
AVLDMYIPQQAQEAATAALLYGDVNPSPGLTQFPFAENQHAVAGDPTSYFQVNDQQT
YRGIHVGRWFDKENVKPLPFHGLSVTSFTSQAPTVPFTSTGGLKAVTVTVRNSGK
RAGQEVVQAYLGASPNVTAPOAKKLVGYTKVSLAAGEAKTVTVNVDRRLQGTGSSA
DLRGATVNVW"
gene
complement(6143..7282)
/gene="desV"
CDS
complement(6143..7282)
/gene="desV"
/codon_start=1
/transl_table=11
/product="putative tautomerase"
/protein_id="AAC68680.1"
/db_xref="GI:3789897"
/translation="MSSRAETPRVPLDLKAAVEELRAETDAALIARVLDGSRVLLGPE
LEGFEAFPAAYCTDRAGVNSGMDALQLALRGLGPGDEVIYPSHTYIASWLAVSA
TGATVPVPEHEDHPTLDPLLVEKAITPRTRALLPVHLXGHPADMALRELADRGLH
IVDAQAAGIARYGRRIAGSAAAFSPYGNKNGCGDGGAVVGTGDEPELAERLRL
RNGSQKYSHETKNTSLRDEMOAAVIRLHAHLDSMNGRRSALAAEYLSLGLGPG
IGLPVAPOTDPMVHLFTVTRRDELRSHLDRGIDTLTHYVPVPHLSPAYAGEAPP
EGSLPRAESFAROVLSLPTGPHLERPQALRVIDADREAEVDAQ"
gene
complement(7279..8292)
/gene="desIV"
CDS
complement(7279..8292)
/gene="desIV"
/codon_start=1
/transl_table=11
/product="TDP-glucose-4,6-dehydratase"
/protein_id="AAC68681.1"
/db_xref="GI:3789898"
/translation="MRLVLTGGAGFTGSHFVROLLAGAYPDVPADEVIVLDSLTVAGN
RANLAPVADPRLRFRVHGDIRDAGILLARELVDAIVHFAAESHVDRSILAGASVETET
NVGOTTLQCAVDAGVGRVHVSTDEVGSDGWTSSPLEPNSPYAASKAGSDL
VARAYHRTGLVDVRLTCCNNYGPQHPKLIPLFVTNLLDGGTLPYDGDANREWV
HTDHCNGIALVLAGRAGEIHHGGLELTHRELGLILLDSLGADWSSYRVKVAURKG
HDLRYSLDGGIEIERLGPQVSVFADGLARTVWYRENNGWWEPLAKATAPQAPATAVE
VSA"
gene
complement(8424..9302)
/gene="desIII"
CDS
complement(8424..9302)
/gene="desIII"
/codon_start=1
/transl_table=11
/product="alpha-D-glucose-1-phosphate
thymidyltransferase"
/protein_id="AAC68682.1"
/db_xref="GI:3789899"
/translation="MKGIVLAGSGTRLHPATSVISKOILPVYKPMIYPIPLVLMLG
GIREQIITPQHIELFQSLGNGHGLIEDYAVQKEPAGIADALLGAETHGDDTC
ALIGDNIFPGPLYTLRDSIARLDGCVLFGYPKDPERYGVAEYDAGRLTDLVEK
PVPKSNLAVTGLYLDNDVDVIAKNIKRPSPRGELEITDVRVYLRGRAELVNLGRG
FAMLDCTDSSLRAAQYQVLEERGVIAGLEETIRMGFIDAEACHGLGEGLSRT
EYGSYLMETIAGREGAP"
gene
complement(9495..10952)
/gene="desII"
CDS
complement(9495..10952)
/gene="desII"
/codon_start=1
/transl_table=11
/product="putative reductase"
/protein_id="AAC68683.1"
/db_xref="GI:3789900"
/translation="MTAPALSATAPAEACPAHGDALGAHVAVGQTLAAGGLVPPDEA
GTTARHILVAVRNGSPPTPLEARHDLGVDRAFRRLALFGVGPPELRTAVETGPA
GAYKNTLLLEORGVDALARKVPYVSVGLYPGTCMFCRCVRYVTGARYDPSA
LDAGNMFVRVIDEIPAGNPSAMFSGGLEPLNPGLSLAHATDHGLRPVYTNFSF
ALERTLEFOPGLWGLHARTISLYGLNDEYEQTGKKAARVRVRNRRFQOLRAER
ESTINGFAYVILPGRASRLDLVDFIADNDAGQGTIDFVNIREDYSGDDKLPQ
EEARLEALNAFEERVRETPGLHDYGVALNSRLTGADALLRKPATMRPTAHPQ
VAVQDLGLGVLYREAGFPDGLATRYIAGRVTPDTSLTEVVRDVERGEVAAVDG
DEYFMGFDQVTVARLNQLERDAADGWEEARGFLR"
gene
complement(10949..12196)
/gene="desI"
CDS
complement(10949..12196)
/gene="desI"
/codon_start=1
/transl_table=11
/product="4-dehydratase"
/protein_id="AAC68684.1"
/db_xref="GI:3789901"
/translation="MKSALSDLAFFGSGPAAFDQPLLVGPNRIDRARLYERLDALDS
QWLSNGGPLVREFEERVAGLAVRHAATCNATAGQLLAHAAGLGEVIMPSMTFAA
TPHALRWIGLTPVFADIDPDTGNLDDPQVAAAVTPRTSAVVGVHLMGRPCAQDLARKV

```





/note="L6290.2; L6910.1L; predicted using Glimmer,  
Testcode and CodonUsage"  
complement(22336..28971)  
/gene="HEL2"  
/note="blastp similarity to helicasess"  
/codon\_start=1  
/product="Helicase 2"  
/protein\_id="AAF35923.1"  
/db\_xref="GI:7025824"

CDS

/translation="MSSHYRELNLQLEVDANGQPDATVETLSASLPKIKOLFQERSLLC  
PSSSEATASNASPTVPVHETALLTCNVLIRWAAERAGQDQGAAGAAARAA  
AVQSDLSQSDQDECRMAKFLATFQAPALVCIDKALTEVAVDFHYAARQFDMF  
VHLSAMERFFIDGSLMLAALSSPHVDMLQMLHVLHNAQVLLRGLQSGRHFHIF  
FRTAWFEPPTQKLTREALRFTAVAAATNPSKLVVDTPFSEFAAAYKVE  
PFLMSDGEOLGHEHLMFNHTEBELLRFEKEGGIGEGTSPASASAPLTDDE  
KLELQRTEDQTMELKRPYSLVDSHGDTVAFYHCFLTWATARRIKVAYSRII  
TRENVAVFVVSFGFARSLVLEPNSTKLAALERAVPLTUSLASLAKTDQSLS  
REQVVSAAHAYLHAGDSPTPEEQQLQCVWVTAICTALVSALRAQSVTSVP  
VAATLNAQPHQCDACATGTEFDVYDGHLLITIMAHLLRTTPAEQLLDEGI  
TDVNLMSQDVVYDGSALITGGALSALPDVILAEPKAMKYPHLTHDILQILARFVE  
GHVSNRPYPATLGNANELGWIIVSOPFDELNDVDVAVVERESQRTVTERERKOLAME  
AAFTVDAQAQAESGIRGFTAGTLAVVCSDESDEDAADTGAAGTSAGKVAKYKQ  
HSGKNNKAREDDIIREANURDANKTVQDWTQVQALCKTKDQASLTVDVADSLGL  
LTAATIRMTGTSFGRNDPGEVLKDSAPVQLQTMWRLVLDASGMREAEAFITDDDD  
ADRSAPADKRRKSAKQATNLYAFGLADIENITNAKSGSHGSDMGKLERLRK  
VKDFTVVDATSYLNLWLYLYVOLHFQCKLMTVVRLQLMQWNERARQAQAPNI  
NVGVPLEYCHHKVLACIRONGVAIISDLDLAVRSALAHDFPASYEYKVDGAIAKW  
NSASVLPSPHPRQKACFTPEMOLMTMGHLERPVIPTKDYRVFNPDCWORELL  
DIVNRGSAVCAPISAGKFTTSYCHWSALKTSNKVVVYVAPNRLINQAVADCA  
RYGSKYSFGRNVGHGADHYRVVDSQVLTLPVLETLILLSPKYNKAKRLDY  
LIDTHIEMSSGNDYMERVALLPCPEVALSATLGTQQLCGWLNVRQLEAQOP  
EAAKDFTHVIDPSSGISORWINDIKKMYLPPGYPQLKTKAPYNNRYIHDLPLS  
ILTLQLOSGPPDITVPEVSVOLYOKMCTLFTFHVFCADGKVAIVQAMKAQSLML  
PETYQOIRYITQOARQYADKNAFAYVNLKGCDCADGLETLSEGEVAENTDMR  
ELICETILQSQRLEDEAOLDKYDAALKPEGRPDTSAEAETASTATAATTA  
IAQPKAFPEKSEFIADNLSVIRELSNRDMPITVTFEGEDCDLTKAILARLED  
AAAYQTEEPKATMERKAAAEARREOSTLKQKLTGDDGEGVADNDVDN  
NNDEDYVDVLPVPEFTFGQYCTVDTPVKEAVEECIKRGDTLCARAIQRGVHH  
AGVKGKLRMEILFRGHGVCATETLALGHSPCRSVVLADGHLNLTITQFQRM  
GRAGRGDLGLHILFSLVSKKIKITLMTSSMTVILKGNVOVDPIOLRLQVLYARH  
RLDKAESWRSLGALMERLYVNPFFAGRAMENGMSFTVEFAQMLLGLFHREG  
HFKDRPSSGLLDVDMVPRNAQGVGFATISMITRKFVAKARLRERYLAPSD  
AVYQAKALEYLLFVSKYVPELHRSALSDPAVIELMSRRVAKRGVYAPLAD  
VCSPATKPFVSVMFSAFYITLASTLKEPADTRPYMCATKKKHCIFGTSMDP  
LMAKLEAPEAAPADPFAVIGSCGHFVHSDLVTLRLGFCODRMLPLVDLTDGC  
RHDAQVLINACITDFVRCQADISVRSNFRFTELLLELNGLSQSESAYVLNTERLS  
NLSGITRDKYTPLOVLSALYPRERGENVAEDLKFTAGTLGVMDFASQLHELPHI  
DKELAWQVMARIEMRRRK"  
complement(29757..>37969)  
/gene="L505.2"  
/note="L6290.3R; predicted using Glimmer, Testcode and  
CodonUsage; Glimmer predicts start codon at 15410"  
complement(29757..>37969)  
/gene="L505.2"  
/codon\_start=3  
/product="L505.2"  
/protein\_id="AAF35924.1"  
/db\_xref="GI:7025825"

gene

CDS

/translation="SRHRRSRHWEPSASATTLAAADAAPQSAATSSPHVHEDTTA  
TSPAPPTQPSSTLSSGAGKQVRAERVGGVAPRAVASSRDSNGNARTSDRSTQ  
QOQOQSDDDANSVPTATPPQYNDFOCRULVALGEARRPLTSEGHMALPENADAG  
VAVPNVDPSQPLAPRTSAPFPPPTAATGSSRRARMPLVFWAPRLIASFTIHG  
ALICLMSQWELTPNNYLLIILFETLNPAGDGRVSDVDAERHEURFVAAWG  
VDGGSSALAAAYSEERHOLGMEYIRRAHPHLFRYVGLPSESASDRSMPSLPVS  
ESRGGLSAAMPNITAAAYCGGMQVRNHRDMYFDSGAVTATGSDVSPASPHSAQ  
TATTPSLMGFGTGHGSGSALPHVLRKDVSHLNTADALVATTPAPSEPSVSP  
PLCLPTGQOQPRPTAKVPMMLDSDPALLADAEGRVLFQSGVRGVOQAVSTTAA  
GEDKYSAQDSTPTPDATKPRVGVASKYSGSAPFSLPAPLQOQRTPTRTATPPSSA  
STVSFVYMAAERVLRRPALWHVAYATPHPTAASCKVOLLOQVRAELVTPFTLL  
PPPSITSSTGSIIPATSAEESQVTOVTLRAEAGVSCERGAADSAAPLPLEVHL  
IGDAGNVQEARVACHLRNRYVSATARKFSTVNPVLLNAGSGSPSTQRTSLRSL  
QSDAPSPAGDVHCHLPLRGLTLHCRNMFLYGLHVRVFNGLVDSVFGVGLTTEELAS  
VQVSRSHLALMPDDRLLRLQPPYRRGFSRDRGAGRGVAPRSTASTASGAADAARPP

PPPPPPPPPPALFYAQRKECVLVLDSSRVESEDTHYRLMPNFHAAAAAATRGMR  
Query Match 71.5%; Score 18.6; DB 3; Length 37969;  
Best Local Similarity 84.0%; Pred. No. 6.7e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 ccgcctaggcggttcggtgtccggc 25  
||||| || ||||| |||||  
Db 14231 CCGCTGGGGGTTCCGGCTCGGC 14255  
AC005929 41944 bp DNA linear INV 24-FEB-2000  
Leishmania major chromosome 3 clone L6910 strain Friedlin, complete  
sequence.  
AC005929  
VERSION AC005929.5 GI:7025830  
KEYWORDS HTG.  
SOURCE Leishmania major.  
ORGANISM Leishmania major.  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
Leishmania.  
REFERENCE 1 (bases 1 to 41944)  
AUTHORS Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrock,M.,  
Cawthra,J., Marsolini,F., Sunkin,S. and Stuart,K.D.  
TITLE Direct Submission  
JOURNAL Submitted (04-NOV-1998) Seattle Biomedical Research Institution, 4  
Nickerson Street, Seattle, WA 98109-1651, USA  
REFERENCE 2 (bases 1 to 41944)  
AUTHORS Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrock,M.,  
Cawthra,J., Marsolini,F., Sunkin,S. and Stuart,K.D.  
TITLE Direct Submission  
JOURNAL Submitted (15-NOV-1999) Seattle Biomedical Research Institution, 4  
Nickerson Street, Seattle, WA 98109-1651, USA  
REFERENCE 3 (bases 1 to 41944)  
AUTHORS Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrock,M.,  
Cawthra,J., Marsolini,F., Sunkin,S. and Stuart,K.D.  
TITLE Direct Submission  
JOURNAL Submitted (24-FEB-2000) Seattle Biomedical Research Institution, 4  
Nickerson Street, Seattle, WA 98109-1651, USA  
COMMENT On Feb 24, 2000 this sequence version replaced gi:6425645.  
FEATURES  
Location/Qualifiers  
1. 41944  
/organism="Leishmania major"  
/strain="Friedlin"  
/db\_xref="taxon:5664"  
/chromosome="3"  
/clone="L6910"  
1420..2418  
/gene="QOR"  
/note="L6910.13; predicted using Glimmer, Testcode and  
CodonUsage"  
1420..2418  
/gene="QOR"  
/codon\_start=1  
/product="Quinone Oxidoreductase"  
/protein\_id="AAF35929.1"  
/db\_xref="GI:7025831"  
/translation="MEVIATAHGGEVLAIVRPSHTPDATQLEGQVILVHNAYGVNF  
IDYFUSGLYKKPAMPYIVGEEGAGVKKVAGVPETMLGKRVAVFGAGCTGYAAF  
TVAPASALREVDPGVTDAAEAACQGLTAHYLVDSYPCGPGSTVVVHAAAGTGILL  
VCQAKLRARVIGVCGGAERATLARGADYVIVATPDWAPLVRAAPQGVDAV  
VDVGQATAGSLSVLRPRGMITFGNAGVAPVPSPLSRSAGSVYLQRPITLGDNR  
TPEVQRTSEVFEFWASKQVQLTIGHEYPLHEAAQALADLSRKYTKLLKCID"  
2880..3416  
/gene="L6910.1"  
/note="predicted using Glimmer, Testcode and CodonUsage"  
2880..3416  
/gene="L6910.1"  
/codon\_start=1  
/product="L6910.1"  
/protein\_id="AAF35930.1"

gene  
/db\_xref="GI:7025832"  
/translation="MKLLKFNVDAPHTLADVQLRNQRLWPCRRHGNAVAAKSWTIT  
AASPVSQGYWLLDDSDQPTICETDVLALPETAANAAGSSGTFAFRGYTKDQVGI  
DFRSKMGQMERATYATGATSKTAKSARTMNAEEMTEKKERNMKREAAEAIED  
NDASVMPKRRKERSKK"  
5389..9468  
/gene="HELI"  
/note="predicted using Glimmer, Testcode and CodonUsage;  
ORF starts at 5110, Glimmer predicts start  
codon at 5389"  
CDS  
5389..9468  
/gene="HELI"  
/note="Blastp similarity to CHL1 helicase"  
/codon\_start=1  
/product="Helicase 1"  
/protein\_id="AAF35931.1"  
/db\_xref="GI:7025833"  
/translation="MLNLSILSYLFDPEDAKVDGTASGSGSVVAEPAAAMATAST  
HDEKAALDYSPPSTDEVLRLQRLDEEMVQVRRRARRVRAQRQIRQARKMLRLQ  
QOLNNGGEHDFLLTODPLAWYAEOLMKKARLDDDLVDLHSASSLSSSSSSSS  
SAREDDDAEAQLEAALSTLPLPKPVFASTRHTQLOLMEDLQRTAFARLPLRR  
HRIEPLAESAGRTAGSGMGVNDTDPATSSKSLGSPHTACPVAAASRPTPSRQQQ  
PRRLTAVHVRPHLCNLAALRRKAGGNDRNLNLYCREAMHFFSKQGRYRQOQEH  
PQYPPHTDRPVRNATEDATGQOELEDDKGCYVCYESHLRSLMTYLRDEQRADAA  
TDPVDAASGASGNGGPPSVYTMRLRLGAELQACPYLATRLLNDGADVAFIPY  
GYLDEQORAVLLGGAATPATAAEAATFSRSTPAAGGGGALONDVAPHSPSL  
GTLYHRQRVATATFRRRRQOQNGVGGDDDEGEAEAIWSTMARSPPSLC  
GDLVFDEANIDACRSTSTVTVPAPHLLARLFTYTLARYASRLTNKQRLREL  
VHFLKSLGSCERADSAVLGEGGEGDIAPLSAPSPMLRLPQPTTPTTAAAT  
LVLPFFHTFDAGIDSDVYAFITFLVDSQLMKLQGFVYALDAELQOQERESYSTAM  
RTMTVSTGDSAGTAGVRKQSGVGVSDAHNEAQRRLASLDLPDNHKKTA  
EGEAPPHLSIAELSQLQAGVAMTAAAHAGAVAPADVOLRALTAELQCVLERL  
CALVSNSTTRVLWTLSPSLSPARPACCAVQKALKVLOLEPGMYTEAPLVEAR  
AVLAGTQMPLAFTCGPILPAQAMWGGDAGCGCTASTIGRSVAVAVGIGSDYV  
HCIDASQGVGAPGPAFAHLISEGHVPSVVQVWALGPGSLKMELSQQALGLR  
SNAASTNDGSSRGIVSSIISPAHRVLAEGCTLLNARVLPAGACICFTSYDM  
DAYVLESTGYTAQISDVKRIFKETRNGGGGGGGGRTAGTGGGGGAEIAEL  
LREYQWISGEPASHAAERAPGSTTASSPVSTAASQOROKPSRRGAFLEAVMG  
RLSEINFADLDGRVAVLGNVPYANPTDVEWMNLKHIVATRLANADARMGRTA  
GSASTPSASSPFTCAEWGLYMDSMRNVQICIGRCIRHAGDYATILLDAYTER  
PDARRRVANLQPSMRVAQTGQCFCGVRFFEAERQPKG"  
10078..11295  
/gene="PAMT"  
/note="L6910.10; predicted using Glimmer, Testcode and  
CodonUsage"  
10078..11295  
/gene="PAMT"  
/codon\_start=1  
/product="Protein Arginine Methyltransferase"  
/protein\_id="AAF35932.1"  
/db\_xref="GI:7025834"  
/translation="MPANRDRFGAPQQRGERSANASKVLTESLATDAENKPTSK  
SLYGDAAARLSSINIMTHDQRLRAYEYIFRNVRKTLHLHGGCMGLYTLMAARGMAK  
MVLGVSSALVDAAVAEQNGKLNILFIRGLCDVLHLPQGMKFYDLCWNGPPLL  
LNFRVLTALYAAHLHTPSGALCPNRAHLVAVSDYSFRLOTEDFNSVNGFOMEP  
MKELVREYEMCALPGSNIVSAPCLAHVTMDTLEGLTVEETATYEAQNAASRLN  
EENPVHRVPTPAVARGVYEAFTLSITRNATVHYLTFTVLDAAFTSKTIPGANFYLA  
VRGGENNTEVSVGLREPLVPVNGEIQGTILRIYTPADKGGKITVYEVTAKEQVVA  
AVETSGYIYQSY"  
12821..13870  
/gene="L6910.11"  
/note="predicted using Glimmer, Testcode and CodonUsage"  
12821..13870  
/gene="L6910.11"  
/codon\_start=1  
/product="L6910.11"  
/protein\_id="AAF35933.1"  
/db\_xref="GI:7025835"  
/translation="MSRVRPHIQRISVEDIMKGNRRLAKAITLAESSNPADSTR  
LGNLLRIHQMATVIRVFPALSGPGSGKSSLDLTIGYLCCKNNHVGVLAVDPSS  
TITHGSLGTCKRKLTHLSYIRFSPSGHLGGVTSRAWEALETFEAREDDVVF  
ETVGVGQSETQCKDLTDMILLVPPASGDELQGIKGVWEVADVWVVTKNDRKKPLV

gene  
/db\_xref="GI:7025832"  
/translation="MKLLKFNVDAPHTLADVQLRNQRLWPCRRHGNAVAAKSWTIT  
AASPVSQGYWLLDDSDQPTICETDVLALPETAANAAGSSGTFAFRGYTKDQVGI  
DFRSKMGQMERATYATGATSKTAKSARTMNAEEMTEKKERNMKREAAEAIED  
NDASVMPKRRKERSKK"  
5389..9468  
/gene="HELI"  
/note="predicted using Glimmer, Testcode and CodonUsage;  
ORF starts at 5110, Glimmer predicts start  
codon at 5389"  
CDS  
5389..9468  
/gene="HELI"  
/note="Blastp similarity to CHL1 helicase"  
/codon\_start=1  
/product="Helicase 1"  
/protein\_id="AAF35931.1"  
/db\_xref="GI:7025833"  
/translation="MLNLSILSYLFDPEDAKVDGTASGSGSVVAEPAAAMATAST  
HDEKAALDYSPPSTDEVLRLQRLDEEMVQVRRRARRVRAQRQIRQARKMLRLQ  
QOLNNGGEHDFLLTODPLAWYAEOLMKKARLDDDLVDLHSASSLSSSSSSSS  
SAREDDDAEAQLEAALSTLPLPKPVFASTRHTQLOLMEDLQRTAFARLPLRR  
HRIEPLAESAGRTAGSGMGVNDTDPATSSKSLGSPHTACPVAAASRPTPSRQQQ  
PRRLTAVHVRPHLCNLAALRRKAGGNDRNLNLYCREAMHFFSKQGRYRQOQEH  
PQYPPHTDRPVRNATEDATGQOELEDDKGCYVCYESHLRSLMTYLRDEQRADAA  
TDPVDAASGASGNGGPPSVYTMRLRLGAELQACPYLATRLLNDGADVAFIPY  
GYLDEQORAVLLGGAATPATAAEAATFSRSTPAAGGGGALONDVAPHSPSL  
GTLYHRQRVATATFRRRRQOQNGVGGDDDEGEAEAIWSTMARSPPSLC  
GDLVFDEANIDACRSTSTVTVPAPHLLARLFTYTLARYASRLTNKQRLREL  
VHFLKSLGSCERADSAVLGEGGEGDIAPLSAPSPMLRLPQPTTPTTAAAT  
LVLPFFHTFDAGIDSDVYAFITFLVDSQLMKLQGFVYALDAELQOQERESYSTAM  
RTMTVSTGDSAGTAGVRKQSGVGVSDAHNEAQRRLASLDLPDNHKKTA  
EGEAPPHLSIAELSQLQAGVAMTAAAHAGAVAPADVOLRALTAELQCVLERL  
CALVSNSTTRVLWTLSPSLSPARPACCAVQKALKVLOLEPGMYTEAPLVEAR  
AVLAGTQMPLAFTCGPILPAQAMWGGDAGCGCTASTIGRSVAVAVGIGSDYV  
HCIDASQGVGAPGPAFAHLISEGHVPSVVQVWALGPGSLKMELSQQALGLR  
SNAASTNDGSSRGIVSSIISPAHRVLAEGCTLLNARVLPAGACICFTSYDM  
DAYVLESTGYTAQISDVKRIFKETRNGGGGGGGGRTAGTGGGGGAEIAEL  
LREYQWISGEPASHAAERAPGSTTASSPVSTAASQOROKPSRRGAFLEAVMG  
RLSEINFADLDGRVAVLGNVPYANPTDVEWMNLKHIVATRLANADARMGRTA  
GSASTPSASSPFTCAEWGLYMDSMRNVQICIGRCIRHAGDYATILLDAYTER  
PDARRRVANLQPSMRVAQTGQCFCGVRFFEAERQPKG"  
10078..11295  
/gene="PAMT"  
/note="L6910.10; predicted using Glimmer, Testcode and  
CodonUsage"  
10078..11295  
/gene="PAMT"  
/codon\_start=1  
/product="Protein Arginine Methyltransferase"  
/protein\_id="AAF35932.1"  
/db\_xref="GI:7025834"  
/translation="MPANRDRFGAPQQRGERSANASKVLTESLATDAENKPTSK  
SLYGDAAARLSSINIMTHDQRLRAYEYIFRNVRKTLHLHGGCMGLYTLMAARGMAK  
MVLGVSSALVDAAVAEQNGKLNILFIRGLCDVLHLPQGMKFYDLCWNGPPLL  
LNFRVLTALYAAHLHTPSGALCPNRAHLVAVSDYSFRLOTEDFNSVNGFOMEP  
MKELVREYEMCALPGSNIVSAPCLAHVTMDTLEGLTVEETATYEAQNAASRLN  
EENPVHRVPTPAVARGVYEAFTLSITRNATVHYLTFTVLDAAFTSKTIPGANFYLA  
VRGGENNTEVSVGLREPLVPVNGEIQGTILRIYTPADKGGKITVYEVTAKEQVVA  
AVETSGYIYQSY"  
12821..13870  
/gene="L6910.11"  
/note="predicted using Glimmer, Testcode and CodonUsage"  
12821..13870  
/gene="L6910.11"  
/codon\_start=1  
/product="L6910.11"  
/protein\_id="AAF35933.1"  
/db\_xref="GI:7025835"  
/translation="MSRVRPHIQRISVEDIMKGNRRLAKAITLAESSNPADSTR  
LGNLLRIHQMATVIRVFPALSGPGSGKSSLDLTIGYLCCKNNHVGVLAVDPSS  
TITHGSLGTCKRKLTHLSYIRFSPSGHLGGVTSRAWEALETFEAREDDVVF  
ETVGVGQSETQCKDLTDMILLVPPASGDELQGIKGVWEVADVWVVTKNDRKKPLV

Query Match 71.5%; Score 18.6; DB 3; Length 41944;  
Best Local Similarity 84.0%; Pred. No. 6.7e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 cccgtctaggcggttcggtccgc 25

```

||||||| !! ||||||| |||||||
Db 38425 CCCGTCGGGGGTTCCGGCTCCGC 38449

RESULT 7
CFU96127/c 1432 bp mRNA linear MAM 28-JAN-1999
LOCUS
DEFINITION Canis familiaris adrenomedullin precursor, mRNA, complete cds.
ACCESSION U96127
VERSION U96127.1 GI:4193235
KEYWORDS
SOURCE
dog,
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 1432)
AUTHORS Ono,Y., Kojima,M., Okada,K. and Kangawa,K.
TITLE cDNA cloning of canine adrenomedullin and its gene expression in
the heart and blood vessels in endotoxin shock
JOURNAL Shock 10 (4), 243-247 (1998)
MEDLINE 99002704
REFERENCE 2 (bases 1 to 1432)
AUTHORS Ono,Y., Kojima,M., Okada,K. and Kangawa,K.
TITLE Direct Submission
JOURNAL Submitted (02-APR-1997) Biochemistry, National Cardiovascular
Center, Fujishiroda 5-7-1, Suita, Osaka 565, Japan
LOCATION/Qualifiers
FEATURES
source
1. .1432
/organism="Canis familiaris"
/db_xref="taxon:9615"
93..155
sig_peptide
93..659
CDS
/codon_start=1
/product="adrenomedullin precursor"
/protein_id="AAD09957.1"
/db_xref="GI:4193236"
/translation="MKLVVALLVGLSLAFGLADTARLDVASEFRKKKKWAVSRGR
EURVSSSYPTGLAEVKGAPQATLIRTDQVKGASRNPTSGPDAARIRVKRYRQSMNF
QGRPSGRCFGTCTVOKLAHQIYQFTDKDGVAPRSKISPGQYGRRRRSLPEPGLR
RTLLEPPRPGGAPAPRAHQVLANLLKM"
156..215
misc_feature
/note="encodes proadrenomedullin N-terminal 20 peptide"
375..530
mat_peptide
/product="adrenomedullin"
BASE COUNT 306 a 426 c 401 g 299 t
ORIGIN

Query Match 70.0%; Score 18.2; DB 4; Length 1432;
Best Local Similarity 87.0%; Pred. No. 1.1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cccgtcaggcttcggttcgcg 23
| ||||| ||| ||||| |||||
Db 170 CACGCTAGCGGTGCGGTGCCG 148

RESULT 8
AF045773/c 1466 bp mRNA linear MAM 20-JAN-1999
LOCUS
DEFINITION Canis familiaris adrenomedullin precursor, mRNA, complete cds.
ACCESSION AF045773
VERSION AF045773.1 GI:4164449
KEYWORDS
SOURCE
dog,
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 1466)
AUTHORS Jougasaki,M., Schirger,J.A., Simari,R.D. and Burnett,J.C. Jr.
TITLE Autocrine role for the endothelin-B receptor in the secretion of
adrenomedullin
JOURNAL Hypertension 32 (5), 917-922 (1998)

```

```

99040162
2 (bases 1 to 1466)
IMOTO,I. and Jougasaki,M.
Cloning of cDNA encoding canine adrenomedullin
Unpublished
REFERENCE 3 (bases 1 to 1466)
AUTHORS Imoto,I. and Jougasaki,M.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-1998) Molecular Cytogenetics, Mayo Clinic and
Foundation, 200 First Street, SW, Rochester, MN 55905, USA
REFERENCE 4 (bases 1 to 1466)
AUTHORS Imoto,I. and Jougasaki,M.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-1999) Molecular Cytogenetics, Mayo Clinic and
Foundation, 200 First Street, SW, Rochester, MN 55905, USA
REMARK Sequence update by submitter
COMMENT On Jan 20, 1999 this sequence version replaced gi:3372495.
FEATURES
Location/Qualifiers
1. .1466
/organism="Canis familiaris"
/db_xref="taxon:9615"
148..714
CDS
/codon_start=1
/product="adrenomedullin precursor"
/protein_id="AAD05423.1"
/db_xref="GI:4164450"
/translation="MKLVVALLVGLSLAFGLADTARLDVASEFRKKKKWAVSRGR
ELRVSSSYPTGLAEVKGAPQATLIRTDQVKGASRNPTSGPDAARIRVKRYRQSMNF
QGRPSGRCFGTCTVOKLAHQIYQFTDKDGVAPRSKISPGQYGRRRRSLPEPGLR
RTLLEPPRPGGAPAPRAHQVLANLLKM"
308 a 441 c 411 g 306 t
BASE COUNT
ORIGIN

Query Match 70.0%; Score 18.2; DB 4; Length 1466;
Best Local Similarity 87.0%; Pred. No. 1.1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cccgtcaggcttcggttcgcg 23
| ||||| ||| ||||| |||||
Db 225 CACGCTAGCGGTGCGGTGCCG 203

RESULT 9
AF147704
LOCUS
DEFINITION
Streptomyces fradiae NDP-hexose 2,3-enoyl reductase TylCII
(tylCII), NDP-hexose 4-ketoreductase TylCIV (tylCIV), NDP-hexose
3-C-methyltransferase TylCIII (tylCIII), mycarosyl transferase
TylCV (tylCV), and NDP-hexose 3,5- (or5-) epimerase TylCVII
(tylCVII) genes, complete cds.
AF147704
ACCESSION
VERSION AF147704.1 GI:5305790
KEYWORDS
SOURCE
Streptomyces fradiae.
ORGANISM Streptomyces fradiae
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 5908)
AUTHORS Bate,N., Butler,A.R., Smith,I.P. and Cundliffe,E.
TITLE The mycarose-biosynthetic genes of Streptomyces fradiae, producer
of tylosin
JOURNAL Microbiology 146 (Pt 1), 139-146 (2000)
MEDLINE 10658660
PUBMED 10658660
REFERENCE 2 (bases 1 to 5908)
AUTHORS Bate,N. and Cundliffe,E.
TITLE Direct Submission
JOURNAL Submitted (30-APR-1999) Biochemistry, University of Leicester,
University Road, Leicester LE1 7RH, UK
LOCATION/Qualifiers
1. .5908
/organism="Streptomyces fradiae"

```

|      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| gene | <pre>/strain="T59235" /db_xref="taxon:1906" complement(80..1069) /gene="tylCII" /note="orf10*" complement(80..1069) /gene="tylCII" /function="biosynthesis of NDP-mycarose during tylosin production" /codon_start=1 /transl_table=11 /product="NDP-hexose 2,3-enoyl reductase TylCII" /protein_id="AAD41821.1" /db_xref="GI:5305791" translation="MSGMYVQLGRGATVLSRLWLTGTVNFGSRVDDQALRLMDEALDR GNCVTDADYGRWLYGKTEELVGRQLSGRGRDDVLAIVGEPMSDRVDRGLSA RHVIRCEASRLYDHDLDLQMHMRDRTVWDELQAMQDLVAGKVRVIGSSNFA GWLHAGQSAARGLSLVSEQLYNLAIVRAELEVPAARAYGIVGPAWPLHGG LSGALRLKLAETAGKSGOQAQRTLPALRDTIARYERFCARVGRDPAEYGLAWLISRP GVSGAVIGPRTTGHLVSAALRAVELESEEHRELEALFPVSGSGVEPWAQON" complement(1066..2109) /gene="tylCIV" /note="orf9*" complement(1066..2109) /gene="tylCIV" /function="biosynthesis of NDP-mycarose during tylosin production" /codon_start=1 /transl_table=11 /product="NDP-hexose 4-ketoreductase TylCIV" /protein_id="AAD41822.1" /db_xref="GI:5305792" translation="MSGMLVPGVGRDRRLVLLGASGFLGSAVAALARPVRLRL VARRSPVAGVAEIEVRRDLISLPGVEAVEADAVVHLVHTGEGKSKRAAGER SEHNVGLRELAEALRPTGPADATRAPVYVLSGLQACMEQAHTPGTYAOKLAER VLRDAEAGVAVVRLRTVIGRSLTSPGSGVIAVTAGRALAGDPTIWHDDGSE RLDLDVDAETFALEHADOLGRHVVYVGTGRHRLDRVFTGVAALAAEHTGRPPV PVYVDPGVAEYCDPTDPSAFRAVTVGRPAEPADGIAAATAAVAGADSGPEP EGAGKRGGGTERR" complement(2106..3362) /gene="tylCIII" /note="orf8*" complement(2106..3362) /gene="tylCIII" /function="biosynthesis of NDP-mycarose during tylosin production" /codon_start=1 /transl_table=11 /product="NDP-hexose 3-C-methyltransferase TylCIII" /protein_id="AAD41823.1" /db_xref="GI:5305793" translation="MFAVPREDQMIISACVCGNRELLPVLDLGEQALGVFPTRREE TVPSIFLELVKSPCGGLVOLRHTPDGLMYGEGVSGIRPFWIGHLRKVAAIR EVLVDLFDLIDGNSDAILKAYPEGRLVIDPSGDKFRELIPPHAEILAEYFSR DVTAFRGVTRARVITSIAMFYDLDPDLAFMRDHDVLDADGIWNEOSYLPAMLEAD ADYVCHHEHLYALQOIEWMAERAGLTIVRAELTDYVGGSLCVTLARASSPHRDEA GPARIARTEAKINTMAPFEFARVVEHQDARDFLDRSRAAGRLTGYCASKGN VILQYCGIGERDLPICIGVSPKAGRTGTGIPIVSEEDAKAMPDQLLVLPWYRE GFVERERDLAGGRLVFLPLRDVV" 3575..4960 /gene="tylCV" /note="orf7*" 3575..4960 /gene="tylCV" /function="catalyses the incorporation of mycarose during tylosin production" /note="glycosyltransferase" /codon_start=1 /transl_table=11 /product="mycarosyl transferase TylCV" /protein_id="AAD41824.1" /db_xref="GI:5305794" translation="MAGLRPGAGVPPTGTPWIPSPGKHCVRSISRRLPYHAPGTP</pre> |
| CDS  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| gene | <pre>GRFRRAPEGSCRMHAIAFFTLPAAGHVNPPTLGVAAELAARGHRTVYALPEDMADRVR RVGARAVTYPLDRERADMPKESDEYTDGEFELKVLWLLDTTADTLPLESFAE EDRDVANDSTFTWTCGLLACKWDIPVIRSTPSYASNEHVALHPPEPGAQVDPAL IELTRAELKLEHGETSDPVAFAATVOSGGLFYMPRIYQYAGETDDRRHFFVGPCA PRASFHTQWQPEDGRPLVMSLGTIYNERGIFRACVAFRDPWMLILVLGGGLGA GDLGPLPENVRDVFPLGVDLPHDTLLVNHGGTSTAMEALAHGVPVIVAMPPEPRA TARRIAELDLGMLLPGEVTAELSGIAQAVLTDDTRIKGLDRMRGEIRRAGGPAVAA DVIIEGLSPAA" 5004..5597 /gene="tylCVII" /note="orf6*" 5004..5597 /gene="tylCVII" /function="biosynthesis of NDP-mycarose during tylosin production" /codon_start=1 /transl_table=11 /product="NDP-hexose 3,5- (orf5-) eplmerase TylCVII" /protein_id="AAD41825.1" /db_xref="GI:5305795" translation="MILITETRVRDAYRITPEPIPHGRGSLVESLRYETLRAUTGHATE IRQVNHVTNRRTNLRGHTVTPGQCKIVTCVGAARTMVVDLVRVSGPFGGHDVVG QDAESGVAVYLPDGLGIGVVALADDTCMNYLTREYTPGMIIIDALDPLGLPWNLT EPPVRSERDAAPSLAEAAAGTLPGEVQCLRAYPAM" 853 a 2123 c 2078 g 854 t BASE COUNT ORIGIN</pre>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| CDS  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| gene | <pre>Query Match 70.0%; Score 18.2; DB 1; Length 5908; Best Local Similarity 87.0%; Pred. NO. 1.le=03; Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0; QY 4 gtcaggcgttcgggtccggcc 26       Db 4514 GTCGAGCGGTCCTCGTACCGGCC 4536       RESULT 10 SCF43 LOCUS Streptomyces coelicolor cosmid F43. ACCESSION ALI36502 VERSION ALI36502.1 GI:6714724 KEYWORDS ABC transporter permease; ABC transporter, ATP-binding component; blaA regulation; cytochrome P450; hydrolase; integral membrane protein; LacI-family transcriptional regulatory protein; prolyl aminopeptidase; RNA polymerase sigma factor; ROK (Nagc/YxLk) family transcriptional regulator; secreted solute-binding lipoprotein; Tetr-family transcriptional regulatory protein; TTA codon. Streptomyces coelicolor A3(2). Streptomyces coelicolor A3(2). Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces. 1 (bases 1 to 22396) Redenbach,M., Kieser,H.M., Denapalte,D., Eichner,A., Cullum,J., Kinashi,H. and Hopwood,D.A. A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome Mol. Microbiol. 21 (1), 77-96 (1996) 97000351 MEDLINE 2 (bases 1 to 22396) REFERENCE Saunders,D.C. and Harris,D. AUTHORS Unpublished JOURNAL REFERENCE 3 (bases 1 to 22396) AUTHORS Thomson,N.R., Parkhill,J., Barrell,B.G. and Rajandream,M.A. TITLE Direct Submission JOURNAL Submitted (17-JAN-2000) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK Notes: Streptomyces coelicolor sequencing at The Sanger Centre is funded</pre>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| CDS  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |

by the BBSRC and Beowulf Genomics  
 Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
 (URL: [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/))  
 CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).  
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.  
 The length in codons is given for each CDS.  
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database where these have been used to deduce the initiation codon.  
 The position of possible ribosome binding site sequences are given using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>.  
**CAUTION:** We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
**IMPORTANT:** This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

# FEATURES

Location/Qualifiers  
 1. .22396  
 /organism="Streptomyces coelicolor A3(2)"  
 /strain="A3(2)"  
 /db\_xref="taxon:100226"  
 /clone="cosmid F43"  
 1. .298  
 /gene="SCF43.01"  
 <1. .298  
 /note="SCF43.01"  
 /note="SCF43.01, possible hydrolase, partial CDS, len: > 98 aa. Identical over the available sequence to Streptomyces lividans TR:P72465(EMBL:Z70724) substrate hyphae associated protein, Orf2 (218 aa), fasta scores opt: 648 z-score: 826.2 E(): 0.100.0% identity in 98 aa overlap. Also similar to several phosphatases e.g. Streptomyces collinus TR:O9X654(EMBL:AF131877) Naph phosphatase (231 aa), fasta scores opt: 186 z-score: 245.7 E(): 2.6e-06 41.9% identity in 86 aa overlap."  
 /codon\_start=2  
 /transl\_table=11  
 /product="putative hydrolase."  
 /protein\_id="CAB66190.1"  
 /db\_xref="GI:6714725"  
 /translation="IDADDALEATASADVDACKPAPPEVHALEAGYPAERAVFG DTWMDRASRGVRCVCGVCGGLPRADLEAGAEAVADPADLLASLSPLA"  
 462. .1601  
 /gene="SCF43.02"  
 462. .1601  
 /note="SCF43.02"  
 /note="SCF43.02, possible integral membrane protein, len: 379 aa. Almost identical to Streptomyces lividans TR:P72466(EMBL:Z70724) substrate hyphae associated protein, Orf3 (329 aa), fasta scores opt: 2115 z-score: 2334.5 E(): 0.93.4% identity in 329 aa overlap. Also weakly similar to Mycobacterium tuberculosis SW:Y08U.MYCTU(EMBL:Z77724) hypothetical 36.8 kD protein (355 aa), fasta scores opt: 208 z-score: 236.0 E(): 9.3e-06 26.1% identity in 303 aa overlap and to a downstream CDS, SCF43.07, putative integral membrane protein (419 aa), fasta scores opt: 276 z-score: 252.1 E(): 1.2e-08 28.1% identity in 366 aa overlap. Note SCF43.02 is 50 aa longer than Streptomyces lividans Orf3.

Contains possible membrane spanning hydrophobic domains."  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative integral membrane protein."  
 /protein\_id="CAB66191.1"  
 /db\_xref="GI:6714726"  
 /translation="MWEACAVGRAVRAWRGPRERDLVQSLKAAGAALIAWTAVG VWLGDPMALMAFWALVLYQATVYSSYRQAGQFRAICTGALLASAAQAINDDNTGAL VLSLPVLMVAVNSRFGGGIYAATAVFVLSGATVSAAGVHGRMGQAALGAVIGVA VNALVFPFLHLDVRENLAAALRAGDLHLSVAADRETQWDAAGMTSDAARLER RLALYSARSWRSRLTSGRLRLRHRAPTVVPEDERWRSVTGNLGAFTFLAY AAEHRTAPAPPGPVLDLYARLLLGIDACHTEAGRWAGERGADPATATETMEELH RLQEGLRHAGQGAARTAVLGLTLLQAEHLWAEIVPEPRESAQ"  
 1887. .1891  
 1898. .2929  
 /gene="SCF43.03"  
 1898. .2929  
 /note="SCF43.03"  
 /note="SCF43.03, unknown, len: 343 aa. Almost identical to Streptomyces lividans TR:P72467(EMBL:Z70724) substrate hyphae associated protein, Orf4 (343 aa), fasta scores opt: 2274 z-score: 2584.7 E(): 0.99.1% identity in 343 aa overlap. Also weakly similar to Streptomyces coelicolor TR:O9WX08(EMBL:AL079345) putative epimerase (353 aa), fasta scores opt: 275 z-score: 318.0 E(): 2.5e-10 25.9% identity in 340 aa overlap."  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein SCF43.03."  
 /protein\_id="CAB66192.1"  
 /db\_xref="GI:6714727"  
 /translation="MSSATGSKIVVTGATGVCTSVVRLLEDPEVGVTLGARIPD WSPAKTEAWADLASDGLTSHFADADAVVHLAWAFQPHDPATPTWNLGIRVFE EAVAAGVPAVHVASQSVGAYSPGKNHADVSEWTHGWPDAAYCKEAYLERALDTFE RDHPGIRVVRMPREFLKRESASORIFRGFLPGAPPELLPFLPDVPLGRLVQAL HTDDAARAYRLAVRSADARGANLAAEPVDAELLGELLGVRLPRAAARSAIAAA WGNLLPASPHLEDAVLRLPVMDCTRAVELGNGATRTATEVLEEFRLGRLQAGADT EPMGRKGVG"  
 complement(2947. .3255)  
 /gene="SCF43.04c"  
 complement(2947. .3255)  
 /note="SCF43.04c"  
 /note="SCF43.04c, unknown, len: 102 aa."  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein SCF43.04c."  
 /protein\_id="CAB66193.1"  
 /db\_xref="GI:6714728"  
 /translation="MSDRHREGPGNGAPVRDMPDQAGDHDHPWEAASPTREQSEH RQGPGERPVDDARSAGDAETGVPDVPDPTDEAGTGRQAGPHSAAVHPEHPVPDESSG"  
 complement(3320. .4609)  
 /gene="SCF43.05c"  
 complement(3320. .4609)  
 /note="SCF43.05c"  
 /note="SCF43.05c, possible ROK (NagC/XylR) family transcriptional regulator, len: 429 aa. Similar to Anaerocellum thermophilum SW:XYLR\_ANA1H(EMBL:Z69782) xylose repressor (399 aa), fasta scores opt: 519 z-score: 575.9 E(): 1.1e-24 28.1% identity in 402 aa overlap. Also highly similar to Streptomyces coelicolor TR:087855(EMBL:AL031013) putative transcriptional regulator SC8A6.21c (441 aa), fasta scores opt: 2034 z-score: 2241.7 E(): 0.71.4% identity in 434 aa overlap. Contains a P1am match to entry PF00480 ROK, rck family and a putative helix-turn-helix motif situated between residues 30. .51 (+3.09 SD)."  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative ROK (NagC/XylR) family transcriptional regulator."  
 /protein\_id="CAB66194.1"  
 /db\_xref="GI:6714729"

/translation="MAGRNGRTVRLRRNRTAVLRQLYFDGSLRFLGPATGLSSG  
SVSNVADIVADGLVEAGSDSDGRPTLLRVAPASGMIQVDVGETRVRFVFLDL  
TLTALAERPLAPDRDVIIVGHVDCGIAEVLATAGLPPERLLGAGTGVPIVEHT  
ADRGAVVHGQTIGWDAVPLEALLRAGSPDPDPVCLIDNGAKTLQGAEMWFGAGGAR  
NAVVLFGSGVSLVTPAEQGRVAVWGHLLTVRVGRRCGALGCLDAYAGASLLI  
ARWREGVRPEGTDEETALTAMLAAYAPDGAADPVALAVLEETAEYVLGAGSLDI  
NLFPORILIGWAGLQGLRFLPVRHRAVSYALRHPARKVTVDLGRIGPDVAIVGA  
AILPLADFFAGRRRPEPEYVPPAWRTALEERAPH"  
complement(3785..4033)  
/gene="SCF43.05c"  
/note="Pfam match to entry PF00480 ROK, ROK family, score  
22.80, E-value 1e-05"  
complement(4614..4618)  
4789..4793  
4799..5293  
/gene="SCF43.06"  
4799..5293  
/gene="SCF43.06"  
/note="SCF43.06"  
protein of unknown function from Streptomyces coelicolor

## misc\_feature

Query Match 70.0%; Score 18.2; DB 1; Length 22396;  
Best Local Similarity 87.0%; Pred. No. 1e+03; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Oy 1 cccgtctaggcgttcggtgtccg 23  
||||| ||||||||| |||||  
Db 12511 CCCGCGAGCGTTCGCTGTCGG 12533

## RESULT 11

AC098934/c  
LOCUS Homo sapiens chromosome 1 clone RP11-480112, WORKING DRAFT  
DEFINITION AC098934 AL390963  
ACCESSION AC098934.1 GI:16751904  
VERSION AC098934.1 GI:16751904  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 173434)  
Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.  
Direct Submission  
Unpublished  
2 (bases 1 to 173434)  
Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.  
Direct Submission  
Submitted (06-NOV-2001) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
On Nov 6, 2001 this sequence version replaced gi:13157584.  
----- Genome Center  
Center: University of Washington Genome Center  
Center Code: UWGC  
Web site: http://www.genome.washington.edu  
Contact: uwgctgs@u.washington.edu  
Drafting Center: SC  
----- Project Information  
Center project name: chr-1  
Center clone name: RP11-480112 (sc0719)  
----- Summary Statistics  
Sequencing vector: plasmid; 32% of reads  
Sequencing chemistry: Dye-terminator ET; 89% of reads  
Chemistry: Dye-terminator Big Dye; 11% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 170284 bases at least Q40  
Consensus quality: 171568 bases at least Q30  
Consensus quality: 172298 bases at least Q20  
Insert size: 172934; sum-of-contrigs  
Quality coverage: 8.7x in Q20 bases; sum-of-contrigs  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 6 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 9114: contig of 9114 bp in length  
\* 9115 9214: gap of unknown length  
\* 9215 27205: contig of 17991 bp in length  
\* 27206 27305: gap of unknown length  
\* 27306 47651: contig of 20346 bp in length  
\* 47652 47751: gap of unknown length  
\* 47752 74121: contig of 26370 bp in length  
\* 74122 74221: gap of unknown length  
\* 74222 112038: contig of 37817 bp in length  
\* 112039 112139: gap of unknown length  
\* 112139 173434: contig of 61296 bp in length.

## FEATURES

## source

1. 173434  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone\_lib="RP11-480112"  
/clone\_lib="RPCI human BAC library 11"  
misc\_feature 1..9114  
/note="assembly\_name:Contig112"  
misc\_feature 9215..27205  
/note="assembly\_name:Contig113"  
misc\_feature 27306..47651  
/note="assembly\_name:Contig114"  
misc\_feature 47752..74121  
/note="assembly\_name:Contig115"  
misc\_feature 74222..112038  
/note="assembly\_name:Contig117"  
misc\_feature 112139..173434  
/note="assembly\_name:Contig118"  
BASE COUNT 47569 a 39082 c 38186 g 48026 t 571 others  
ORIGIN

Query Match 70.0%; Score 18.2; DB 2; Length 173434;  
Best Local Similarity 87.0%; Pred. No. 9,4e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 cgtctaggcgttcggtgtccggc 25  
| ||||| ||||| ||||| ||

Db 152779 CTTCTAGCGTTCAGTGTCTCGC 152757

## RESULT 12

## SLMELOPER/c

LOCUS SLMELOPER 3149 bp DNA linear BCT 21-FEB-1996  
DEFINITION S.lincolnensis melC1, melC2, ORF3 and rnhH genes.  
ACCESSION X95703  
VERSION X95703.1 GI:1200138

KEYWORDS mel-operon; melC1 gene; melC2 gene; ORF3; RNase H; rnhH gene;  
tyrosinase; tyrosinase co-factor.  
Streptomyces lincolnensis.  
Streptomyces lincolnensis.  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

## REFERENCE

1 (bases 1 to 3149)  
Zhang,H.Z. and Piepersberg,W.  
Cloning and characterization of the mel-operon from S.lincolnensis  
78-11  
JOURNAL Unpublished  
2 (bases 1 to 3149)  
Piepersberg,W.  
Direct Submission  
TITLE  
Submitted (15-FEB-1996) W. Piepersberg, Bergische Univ.  
Gesamthochschule, Wuppertal, Institut fuer Chemische Mikrobiologie,



|             |                                                         |            |             |     |        |                 |
|-------------|---------------------------------------------------------|------------|-------------|-----|--------|-----------------|
| FEATURES    | Gausstr. 20, 42097 Wuppertal, FRG                       | AF319543   | 6927 bp     | DNA | linear | BCT 11-JAN-2001 |
| source      | Location/Qualifiers                                     | LOCUS      |             |     |        |                 |
|             | 1. .3149                                                | DEFINITION |             |     |        |                 |
|             | /organism="Streptomyces lincolnensis"                   | ACCSSION   |             |     |        |                 |
|             | /strain="78-11"                                         | VERSION    |             |     |        |                 |
| gene        | /db_xref="taxon:1915"                                   | AF319543.1 | GI:12082814 |     |        |                 |
|             | 10. .432                                                | KEYWORDS   |             |     |        |                 |
| CDS         | /gene="melC1"                                           | SOURCE     |             |     |        |                 |
|             | 10. .432                                                | ORGANISM   |             |     |        |                 |
|             | /codon_start=1                                          |            |             |     |        |                 |
|             | /transl_table=11                                        | REFERENCE  |             |     |        |                 |
|             | /product="Tyrosinase co-factor (MelC2)"                 | AUTHORS    |             |     |        |                 |
|             | /protein_id="CAA64999.1"                                | TITLE      |             |     |        |                 |
|             | /db_xref="GI:1200139"                                   |            |             |     |        |                 |
|             | /db_xref="SWISS-PROT:P55048"                            | JOURNAL    |             |     |        |                 |
|             | /translation="MPRLTERRALTAAMALASSAGAGAGAAACGAAAHGSDPV   | REFERENCE  |             |     |        |                 |
| gene        | P1PCSLLDLPILLDEVIKRGIOGPHRGHGTGYAFLDGVELHVMRAD          | AUTHORS    |             |     |        |                 |
|             | 465. .1286                                              | TITLE      |             |     |        |                 |
|             | GWISVSHYDPVPTPRAAARAADLOGAKLVPPAN"                      | JOURNAL    |             |     |        |                 |
| CDS         | /gene="melC2"                                           | REFERENCE  |             |     |        |                 |
|             | 465. .1286                                              | AUTHORS    |             |     |        |                 |
|             | /gene="melC2"                                           | TITLE      |             |     |        |                 |
|             | /codon_start=1                                          | JOURNAL    |             |     |        |                 |
|             | /transl_table=11                                        | REFERENCE  |             |     |        |                 |
|             | /product="Tyrosinase"                                   | AUTHORS    |             |     |        |                 |
|             | /protein_id="CAA65000.1"                                | TITLE      |             |     |        |                 |
|             | /db_xref="GI:1200140"                                   | JOURNAL    |             |     |        |                 |
|             | /db_xref="SWISS-PROT:P55023"                            | FEATURES   |             |     |        |                 |
|             | /translation="WYRKQNTLTADEXRRVTAVLSSSARVDYFTVTHNEFIVAD  | source     |             |     |        |                 |
| gene        | TUNGERTGRSPFLPWHRRFLLEFALQSDASVAPLYWDWSTDRASLWADPF      |            |             |     |        |                 |
|             | 1369. .1821                                             | CDS        |             |     |        |                 |
| CDS         | LGKTRGRNGVTPGPPRAATGVPITVRLDGRYLRALGAGRELPTPRAEYDVL     |            |             |     |        |                 |
|             | 1369. .1821                                             |            |             |     |        |                 |
|             | IPTYDMAPNNSASDGFNRHLEGRVNLHNRVHVWVGOMATGVSPNDPVFWLHHAYI |            |             |     |        |                 |
|             | 1369. .1821                                             |            |             |     |        |                 |
|             | DKLMAQWRHRTPAYVPAAGTPDVVDLDTMKPNHDSPADLLDHTGHVTFDTD"    |            |             |     |        |                 |
| gene        | /gene="ORF3"                                            |            |             |     |        |                 |
|             | 1369. .1821                                             |            |             |     |        |                 |
| CDS         | /gene="ORF3"                                            |            |             |     |        |                 |
|             | /codon_start=1                                          |            |             |     |        |                 |
|             | /transl_table=11                                        |            |             |     |        |                 |
|             | /protein_id="CAA65001.1"                                |            |             |     |        |                 |
|             | /db_xref="GI:1200141"                                   |            |             |     |        |                 |
|             | /db_xref="SWISS-PROT:P55049"                            |            |             |     |        |                 |
|             | /translation="MSKVESIEGVVHITATNQTPETPPQMSGVERIEQRTDL    |            |             |     |        |                 |
| gene        | THWTSVNGVHKEFDAITEIQIDERVAVTTVGGAEQAGVYTFRLDDDDHTKVMLOM |            |             |     |        |                 |
|             | 1369. .1821                                             |            |             |     |        |                 |
|             | DFHPSVTEKYQDKLGFVKRQTKDLEREKKEIERGQETGGWRGAVI"          |            |             |     |        |                 |
| CDS         | complement(2204. .2557)                                 |            |             |     |        |                 |
|             | complement(2204. .2557)                                 |            |             |     |        |                 |
|             | /gene="rnh"                                             |            |             |     |        |                 |
|             | /codon_start=1                                          |            |             |     |        |                 |
|             | /transl_table=11                                        |            |             |     |        |                 |
|             | /product="putative RNaseH"                              |            |             |     |        |                 |
|             | /protein_id="CAA65002.1"                                |            |             |     |        |                 |
|             | /db_xref="GI:1200142"                                   |            |             |     |        |                 |
|             | /db_xref="SPTREMBL:Q54388"                              |            |             |     |        |                 |
|             | /translation="MRERVAAACDGASKNPGPAGWAWVADDTFTPTWEAGALGKA |            |             |     |        |                 |
|             | TNNVAELERLLAATRDVPLEIRMDSQYANKAVTTWLPQWRNGWKTAAAGKPVAN  |            |             |     |        |                 |
|             | QDLVVRIDELDGRS"                                         |            |             |     |        |                 |
| BASE COUNT  | 480 a 1140 c 1071 g 458 t                               |            |             |     |        |                 |
| ORIGIN      |                                                         |            |             |     |        |                 |
|             | 69.2%; Score 18; DB 1; Length 3149;                     |            |             |     |        |                 |
| Query Match | Best Local Similarity 80.8%; Pred. No. 1.4e+03;         |            |             |     |        |                 |
| Matches     | 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;     |            |             |     |        |                 |
| QY          | 1 cccgtctaggcgttcgggtccggcc 26                          |            |             |     |        |                 |
|             |                                                         |            |             |     |        |                 |
| DB          | 866 CCCGTCAGGCGCACGCGTGTGCGGCC 841                      |            |             |     |        |                 |
| RESULT      | 13                                                      |            |             |     |        |                 |
|             | Query Match 69.2%; Score 18; DB 1; Length 6927;         |            |             |     |        |                 |
|             | Best Local Similarity 80.8%; Pred. No. 1.3e+03;         |            |             |     |        |                 |
|             | Query Match 69.2%; Score 18; DB 1; Length 6927;         |            |             |     |        |                 |
|             | Best Local Similarity 80.8%; Pred. No. 1.3e+03;         |            |             |     |        |                 |



**Qy**      1 cccgtctagcggttcggtgtcgcggcc 26  
         || || ||||| ||||| |||||  
**Db**      3493 CCGCTCCAGGCGGTGGGTCTCCGGCC 3518

SULT 14  
109873  
CUS  
AC109873 20633 bp DNA linear HTG 08-FEB-2002  
Rattus norvegicus clone CH230-303015, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 12 unordered pieces.

| SESSION | AC109873          | AC109873.1 | GI:18640591 |
|---------|-------------------|------------|-------------|
| SESSION | AC109873          | AC109873.1 | GI:18640591 |
| WORDS   | HTG; HTGS_PHASE1, |            |             |
| WORDS   | Norway rat.       |            |             |
| WORDS   | HTG; HTGS_PHASE1, |            |             |
| WORDS   | Norway rat.       |            |             |

| ORGANISM                                                                                                                           | REFERENCE            |
|------------------------------------------------------------------------------------------------------------------------------------|----------------------|
| <i>Rattus norvegicus</i>                                                                                                           | 1 (bases 1 to 20633) |
| Eukaryota; Chordata; Vertebrata; Euteleostomi;<br>Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;<br><i>Rattus</i> . |                      |

AUTHORS

Muzny, D.M., Adams, C., Adio-oduola, B., Ali-osman, F.R., Allen, C., Allbrooks, S.L., Amaratinge, H.C., Are, J.R., Ayelle, M., Banks, T., Barbaria, J., Benton, J., Bimege, K., Blankenburg, K., Bonnik, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davilla, M.L., Davis, C., Davy-Carroll, L., Dederich, D.H., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flaeg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haviak, P., Hawes, A., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homisi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jameson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseghe, H., Lozada, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, S., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwu, S., Oguh, M., Okwundu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.I., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Verd, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G., and Gibbs, R.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS

**TITLE** Direct Submission  
**JOURNAL** Submitted (08-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

MENT  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hqsc.bcm.tmc.edu/>  
----- Houston, TX 77030, USA

```

Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: Q10
Center clone name: CH230-303015
----- Summary Statistics
Sequencing vector: Plasmid; M77789
Chemistry: Dye-terminator Big Dye; 98% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 22789 bases at least Q40
Consensus quality: 25322 bases at least Q30
Consensus quality: 27555 bases at least Q20
Estimated insert size: 16593; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 0.1x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

|       |        |                |              |
|-------|--------|----------------|--------------|
| 1     | 1524:  | contig of 1524 | bp in length |
| 1525  | 1624:  | gap of unknown | length       |
| 1625  | 3410:  | contig of 1786 | bp in length |
| 3411  | 3510:  | gap of unknown | length       |
| 3511  | 4516:  | contig of 1006 | bp in length |
| 4517  | 4616:  | gap of unknown | length       |
| 4617  | 6190:  | contig of 1574 | bp in length |
| 6191  | 6290:  | gap of unknown | length       |
| 6291  | 8737:  | contig of 2447 | bp in length |
| 8738  | 8837:  | gap of unknown | length       |
| 8838  | 10858: | contig of 2021 | bp in length |
| 10859 | 10958: | gap of unknown | length       |
| 10959 | 13211: | contig of 2253 | bp in length |
| 13212 | 13311: | gap of unknown | length       |
| 13312 | 14554: | contig of 1243 | bp in length |
| 14555 | 14654: | gap of unknown | length       |
| 14655 | 16077: | contig of 1423 | bp in length |
| 16078 | 16177: | gap of unknown | length       |
| 16178 | 17711: | contig of 1534 | bp in length |
| 17712 | 17811: | gap of unknown | length       |
| 17812 | 19102: | contig of 1291 | bp in length |
| 19103 | 19202: | gap of unknown | length       |
| 19203 | 20633: | contig of 1431 | bp in length |

FEATURES  
source

BASE COUNT  
ORIGIN

BASE COUNT  
ORIGIN

Query Match 69.2%; Score 18; DB 2; Length 20633;  
Best Local Similarity 80.8%; Pred. No. 1.3e+03;

QY 1 ccggtctaggcggttcggtgtgtccggcc 26  
||||||| | ||||| | ||  
11250 CCCGTCTAGCCCTTCGGTGTGTGCC 11275  
Db

RESULT 15  
LMFL6586/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS

LMFL6586 40064 bp DNA linear INV 20-MAR-2001  
Leishmania major Friedlin chromosome 23 cosmid L6586.  
AL049768  
AL049768.2 GI:6855406  
conserved WD40 repeat domain protein; kinesin-related protein.

transmembrane protein; ubiquitin activating enzyme.  
 Leishmania major.  
 Leishmania major  
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
 Leishmania.  
 1 (bases 1 to 40064)  
 Ivens, A.C., Lewis, S.M., Bagherzadeh, A., Zhang, L., Chan, H.M. and  
 Smith, D.F.  
 A physical map of the Leishmania major Friedlin genome  
 Genome Res. 8 (2), 135-145 (1998)  
 2 (bases 1 to 40064)  
 Klages, S., Borzym, K., Reinhardt, R., Beck, A., Ivens, A.C., Quail, M.,  
 Rajandream, M.A. and Barrell, B.G.  
 Direct Submission  
 Submitted (16-MAR-2001) European Leishmania major Friedlin genome  
 sequencing project, Sanger Centre, The Wellcome Trust Genome  
 Campus, Hinxton, Cambridge CB10 1SA, (E-mail: barrell@sanger.ac.uk)  
 and Max Planck Institut fuer Molekulare Genetik, Ihnestr. 73,  
 D-14195 Berlin, Germany  
 On Feb 2, 2000 this sequence version replaced gi:4760361.  
 See <http://www.ebi.ac.uk/parasites/leish.html>  
 Notes:  
 Details of leishmania sequencing at the Sanger Centre are available  
 on the World Wide Web.  
 see <http://www.sanger.ac.uk/projects/Lmajor/>  
 CDS are numbered using the following system eg L6586.01. L6586  
 (cosmid name), .01 (first CDS)  
 To make the cosmid library Leishmania major Friedlin DNA was  
 partially digested with Sau3AI prior to cloning into BamHI site of  
 the cosmid shuttle vector cLMVG (Ryan et al. 1993 Gene  
 131:145-150). The sequence of the packaged vector was determined by  
 Peter Myler and Ken Stuart at Seattle Biomedical Research  
 Institute, and is available as accession number U59231.  
 The more significant matches with motifs in the PROSITE database  
 are also included but some of these may be fortuitous. The length  
 in codons is given for each CDS.  
 Usually the highest scoring match found by fasta -o is given for  
 CDS which show significant similarity to other CDS in the database.  
 Gene prediction is done using:  
 (1) the FramePlot program of Bibb et al.,  
 Gene 30:157-166(1984) as implemented  
 at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. (2)  
 codon preference based on the codon usage table for Leishmania at  
<http://www.kazusa.or.jp/codon/>  
 (3)  
 the Hexamer program which was written by Richard Durbin as an  
 integral part of the ACEDB-based analysis tools for the C.elegans  
 Genome Sequencing project. The program calculates the  
 log-likelihood score for a given DNA segment based on the frequency  
 of 6-mers, normalised for the base-pair composition of the genome.  
 The program was trained using a fasta file of confirmed Leishmania  
 major coding sequences (CDS), i.e. from AUG start codon to the stop  
 codon.  
 CAUTION: We may not have predicted the correct initiation codon.  
 Where possible we choose an initiation codon (atg) which is  
 preceded by a stretch of pyrimidines or part of a kozak sequence.  
 If this cannot be identified we choose the most upstream initiation  
 codon. Transmembrane domains were predicted as implemented at the  
 TMHMM server: <http://www.cbs.dtu.dk/services/TMHMM-1.0/>  
 IMPORTANT: This sequence MAY NOT be the entire insert of the  
 sequenced clone. It may be shorter because we only sequence  
 overlapping sections once, or longer, because we arrange for a  
 small overlap between neighbouring submissions. Cosmid L6586 is not  
 overlapped by any clones sequenced to date. It contains the left  
 end of PAC P460 (AL160762), and the right end of PAC P986  
 (AL390693).  
 Location/Qualifiers  
 1. 40064  
 /organism="Leishmania major"  
 /strain="Friedlin"

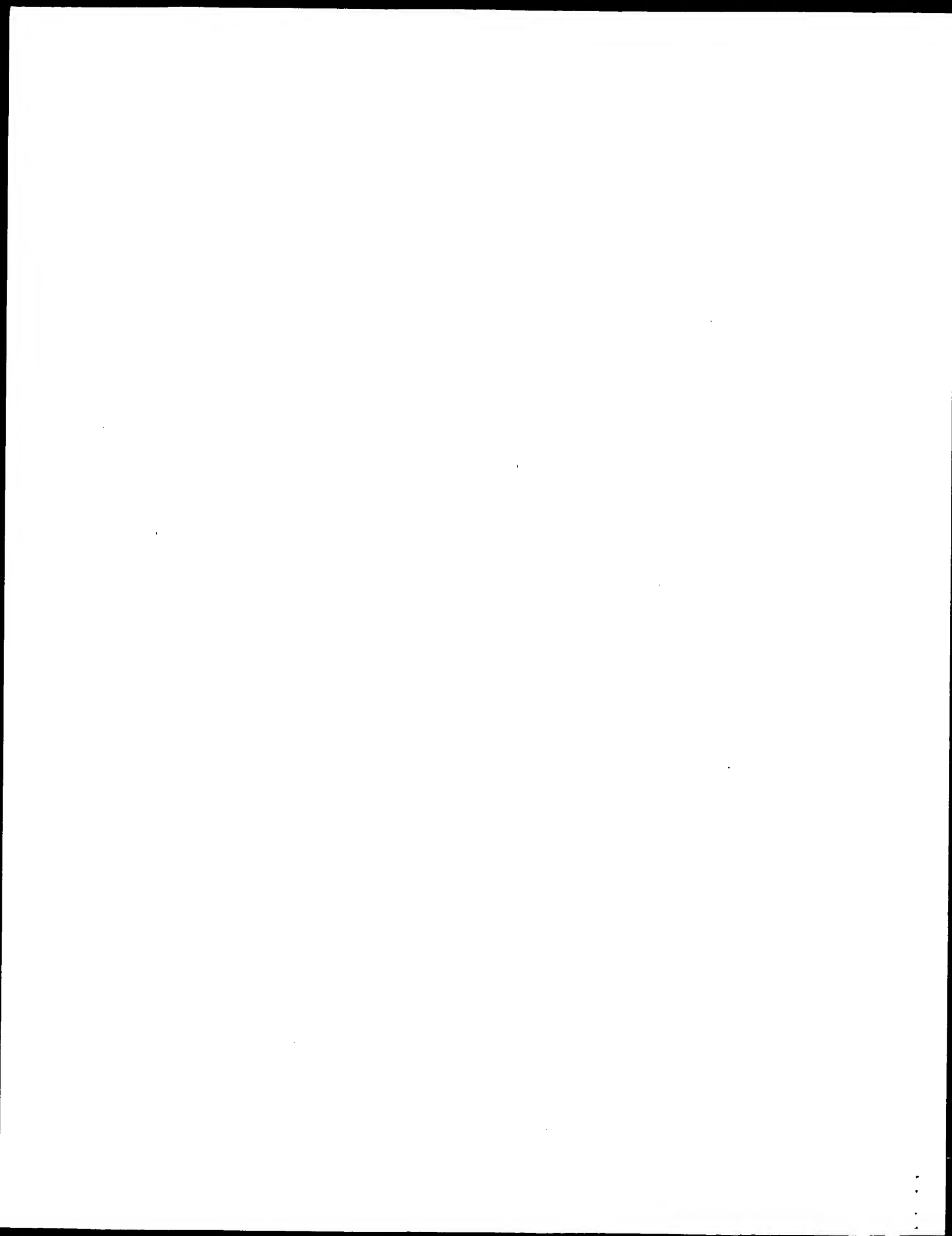
/db\_xref="taxon:5664"  
 /chromosome="23"  
 /clone="cosmid L6586"  
 153..170  
 /note="(gcgtgt)3"  
 488..512  
 /note="poly-pyrimidine tract"  
 1170..1185  
 /note="poly-pyrimidine tract"  
 complement(1314..1707)  
 /note="region of BLASTN similarity to: AFI39112 Leishmania  
 peruviana clone AC16 microsatellite sequence., bases  
 1..394, 83% identity over 393 bases"  
 1337..3379  
 /gene="L6586.01"  
 1337..3379  
 /gene="L6586.01"  
 /note="L6586.01, len = 679 aa, conserved hypothetical  
 protein; predicted pi = 9.8667; predicted coiled-coil  
 region at aa  
 590-610; good similarity to AF253317 Trypanosoma cruzi  
 LY1p (LY1) gene; contains two Pfam matches to entry  
 PF00400 WD40, WD domain, G-beta repeat; reasonable  
 similarity to Q9PS7, conserved hypothetical protein (554  
 aa, Neurospora crassa, EMBL: AL355932, CAB91444); Fasta  
 scores: E():0, 34.9% identity in 502 aa"  
 /codon\_start=1  
 /label="L6586.01"  
 /product="conserved WD40 repeat domain protein"  
 /protein\_id="CAB71230.1"  
 /db\_xref="GI:6855407"  
 /db\_xref="SP:PREMBL:Q9NF84"  
 /translation="MPPTSLTFSRSSPPVATAAPRVGAQRKPSKEVQELIDKRIEEL  
 KQETGQHEKFIQGRVWDTAALNRGRTRRERLRAPKRAKVDMYIOMEEDNIGL  
 LLDHEVADRSIQSDIASQIDQOKRYSLVDKLPYKIDFINSINGTHLLAGLRGH  
 MANLRGKSFQEGCTADKIDRISDAIFLVDSMTAVAKKRVMTYKTEGTHLLSKRA  
 HMDRLGVLPHKMLLAATSTYTMOYLDISTGQEVGTVKPAVMRDPSTCLAVNPSNGV  
 AATCDLGVKVPSPVDPFLQLKGKGVIEDICEHPNGRFLTLGGDHAKMYWDCR  
 TLRTEGVATYTSFHTLIDISSGLVALGGTNVHWKDMPTAAKSPSPYKRGVGN  
 TAEQVRCFPIGVIGHSRGRTSLIIPSGEANPDFYANPHEHRRKRVNTLL  
 DKLPDITISMDIQVPGVNEKRLAENLRLNRKARAIREKMKRASKSLGGAAPTGL  
 LVGRDEEVDEELGYKEAPKTEVKRQELERKRWKDKKDSADKVRKSKQTLRASRI  
 VQRNRAQMDARNGVDFENALDAEETALAAASRAAKQHKKLARKEALEIANDIRGO  
 LRQVDYDLVPPAQRREFEKSYHSGSDTAGRGTEGSSGPEAEARTNAALKRFLR"  
 1424..2232  
 /gene="L6586.01"  
 /note="region of BLASTN similarity to: AF253317  
 Trypanosoma cruzi LY1p (LY1) gene, LY1-a allele,  
 complete cds., bases 4054..4922, 66% identity over 868  
 bases"  
 complement(1578..2185)  
 /note="region of BLASTN similarity to: AQ644878  
 RPI93-DpnII-25G1.TV RPI93-DpnII Trypanosoma brucei  
 genomic clone RPI93-DpnII-25G1, bases 1..608, 67%  
 identity over 607 bases"  
 2222..2332  
 /gene="L6586.01"  
 /note="Pfam match to entry PF00400 WD40, WD domain, G-beta  
 repeat, score 25.60, E-value 0.0012"  
 2335..3112  
 /gene="L6586.01"  
 /note="region of BLASTN similarity to: AF253317  
 Trypanosoma cruzi LY1p (LY1) gene, LY1-a allele,  
 complete cds., bases 4974..5751, 71% identity over 777  
 bases"  
 2348..2449  
 /gene="L6586.01"  
 /note="Pfam match to entry PF00400 WD40, WD domain, G-beta  
 repeat, score 7.00, E-value 66"  
 2630..3063  
 /gene="L6586.01"  
 /note="region of BLASTN similarity to: AL483866 TA292D02P  
 Trypanosoma brucei TREU927 sheared genomic DNA Trypanosoma

```
brucei genomic clone 292d02 forward, bases 1. .434, 67%
identity over 433 bases"
repeat_region 3522. .3542
repeat_region /note="poly-pyrimidine tract"
repeat_region 3908. .3922
repeat_region /note="(ctc)5"
repeat_region 3908. .3926
repeat_region /note="poly-pyrimidine tract"
repeat_region 3995. .4010
repeat_region /note="poly-pyrimidine tract"
repeat_region complement(4052. .4089)
repeat_region /note="poly-pyrimidine tract"
repeat_region 4068. .4085
repeat_region /note="(ggggag)3"
repeat_region 4717. .4739
repeat_region /note="poly-pyrimidine tract"
repeat_region complement(5114. .5131)
repeat_region /note="poly-pyrimidine tract"
repeat_region complement(5359. .5376)
repeat_region /note="poly-pyrimidine tract"
repeat_region 5390. .5405
repeat_region /note="poly-pyrimidine tract"
repeat_region 5561. .5577
repeat_region /note="poly-pyrimidine tract"
misc_feature 5908. .6414
/note="region of BLASTN similarity to: A0847049
LMAJFV1_lm37e03.xl Leishmania major FV1 random genomic
library Leishmania major genomic clone LMAJFV1_lm37e03 3',
bases 1. .507, 100% identity over 506 bases"
misc_feature 6300. .6849
/note="hexamer gene prediction, score 72.0664; previously
part of L6586.02"
misc_feature complement(6837. .7217)
/note="region of BLASTN similarity to: A0850952
LMAJFV1_lm37e03.yl Leishmania major FV1 random genomic
library Leishmania major genomic clone LMAJFV1_lm37e03 5',
bases 1. .183, 99% identity over 182 bases, bases 184. .380,
99% identity over 196 bases"
repeat_region 7815. .7826
repeat_region /note="(gct)4"
repeat_region 8687. .8698
repeat_region /note="(cctc)3"
repeat_region 8845. .8862
repeat_region /note="poly-pyrimidine tract"
repeat_region 8882. .8914
repeat_region /note="poly-pyrimidine tract"
repeat_region 9523. .9539
repeat_region /note="poly-pyrimidine tract"
repeat_region 9951. .9969

Query Match 69.2%; Score 18; DB 3; Length 40064;
Best Local Similarity 80.8%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 ccgctcaggcgttcggtccggcc 26
||||| ||| ||| ||| |||
Db 25046 CCCGCTTGGCATTCTGTCGCGACC 25021
```

Search completed: July 31, 2002, 14:01:53  
Job time: 17555 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:20:35 ; Search time 165.21 Seconds  
(without alignments)  
107.049 Million cell updates/sec

Title: US-09-899-718A-4  
Perfect score: 72  
Sequence: 1 ccttcaggacgatgcttcg.....catgtgagcccaacagtgc 72

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA: \*  
1: /cqn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cqn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cqn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cqn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cqn2\_6/ptodata/2/ina/PCUS\_COMB.seq: \*  
6: /cqn2\_6/ptodata/2/ina/backfiles1.seq: \*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| 1          | 23.8  | 33.1        | 2529   | 2  | US-08-883-795A-37  |
| 2          | 23.2  | 32.2        | 87350  | 3  | US-08-781-891-79   |
| 3          | 23    | 31.9        | 3187   | 5  | PCR-US95-06815-1   |
| 4          | 23    | 31.9        | 3192   | 1  | US-08-706-037-26   |
| 5          | 23    | 31.9        | 3192   | 2  | US-09-005-397-26   |
| 6          | 22.6  | 31.4        | 2375   | 2  | US-08-903-800A-5   |
| 7          | 22.2  | 30.8        | 80161  | 3  | US-09-036-987A-1   |
| 8          | 22.2  | 30.8        | 80161  | 4  | US-09-370-700-1    |
| 9          | 22    | 30.6        | 1968   | 4  | US-09-232-191-30   |
| 10         | 22    | 30.6        | 1968   | 4  | US-09-232-200-80   |
| 11         | 22    | 30.6        | 1968   | 4  | US-09-232-197-80   |
| 12         | 22    | 30.6        | 1968   | 4  | US-09-232-201-80   |
| 13         | 22    | 30.6        | 246240 | 2  | US-08-724-394A-20  |
| 14         | 22    | 30.6        | 246240 | 2  | US-08-724-394A-21  |
| 15         | 22    | 30.6        | 246240 | 2  | US-08-724-394A-22  |
| 16         | 21.8  | 30.3        | 765    | 4  | US-08-998-416-1046 |
| 17         | 21.8  | 30.3        | 876    | 4  | US-09-315-793-61   |
| 18         | 21.8  | 30.3        | 1279   | 3  | US-08-985-950-5    |
| 19         | 21.8  | 30.3        | 1728   | 3  | US-08-985-950-7    |
| 20         | 21.8  | 30.3        | 3057   | 1  | US-08-551-459-3    |
| 21         | 21.8  | 30.3        | 3466   | 1  | US-08-551-459-5    |
| 22         | 21.6  | 30.0        | 107    | 4  | US-09-042-353-251  |
| 23         | 21.6  | 30.0        | 107    | 4  | US-08-758-417A-99  |
| 24         | 21.6  | 30.0        | 108    | 1  | US-08-053-131-134  |
| 25         | 21.6  | 30.0        | 108    | 1  | US-08-645-641-134  |
| 26         | 21.6  | 30.0        | 108    | 1  | US-07-853-408B-134 |
| 27         | 21.6  | 30.0        | 108    | 1  | US-07-096-762-134  |

|    |      |      |      |   |                    |                   |
|----|------|------|------|---|--------------------|-------------------|
| 28 | 21.6 | 30.0 | 108  | 2 | US-08-308-865-134  | Sequence 134, App |
| 29 | 21.6 | 30.0 | 108  | 5 | PCT-US92-10983-134 | Sequence 134, App |
| 30 | 21.6 | 30.0 | 3278 | 1 | US-07-778-890A-1   | Sequence 1, Appli |
| 31 | 21.4 | 29.7 | 2249 | 3 | US-08-814-052-19   | Sequence 19, Appl |
| 32 | 21.4 | 29.7 | 2279 | 3 | US-08-814-052-17   | Sequence 17, Appl |
| 33 | 21.4 | 29.7 | 2300 | 3 | US-08-814-052-18   | Sequence 18, Appl |
| 34 | 21.4 | 29.7 | 3183 | 2 | US-08-939-218A-1   | Sequence 1, Appli |
| 35 | 21.4 | 29.7 | 3192 | 1 | US-08-940-661A-1   | Sequence 1, Appli |
| 36 | 21.4 | 29.7 | 3192 | 2 | US-09-083-485-1    | Sequence 1, Appli |
| 37 | 21.2 | 29.4 | 1174 | 2 | US-08-793-410-8    | Sequence 8, Appli |
| 38 | 21.2 | 29.4 | 1604 | 4 | US-09-438-938-9    | Sequence 9, Appli |
| 39 | 21.2 | 29.4 | 4325 | 1 | US-08-453-924-2    | Sequence 2, Appli |
| 40 | 21.2 | 29.4 | 4325 | 1 | US-08-471-791-29   | Sequence 29, Appl |
| 41 | 21.2 | 29.4 | 4325 | 5 | PCT-US91-01746-29  | Sequence 29, Appl |
| 42 | 21.2 | 29.4 | 9323 | 1 | US-08-038-682-6    | Sequence 6, Appli |
| 43 | 21.2 | 29.4 | 9323 | 1 | US-08-302-832-6    | Sequence 6, Appli |
| 44 | 21.2 | 29.4 | 9323 | 2 | US-08-530-198-6    | Sequence 6, Appli |
| 45 | 21.2 | 29.4 | 9323 | 2 | US-08-469-880-6    | Sequence 6, Appli |

## ALIGNMENTS

RESULT 1  
US-08-883-795A-37  
; Sequence 37, Application US/08883795A  
; Patent No. 5985607  
; GENERAL INFORMATION:  
; APPLICANT: Delcuv, Genevieve  
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BERESKIN & PARR  
; STREET: 40 King Street West  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5H 3Y2  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/883,795A  
; FILING DATE: 27-JUN-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gravelle, Micheline  
; REGISTRATION NUMBER: 40,261  
; REFERENCE/DOCKET NUMBER: 7841-062  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 364-7311  
; TELEFAX: (416) 361-1398  
; INFORMATION FOR SEQ ID NO: 37:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2529 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: Rh 10  
US-08-883-795A-37

Query Match 33.1%; Score 23.8; DB 2; Length 2529;  
Best Local Similarity 59.7%; Pred. No. 2.3; Indels 0; Gaps 0;  
Matches 40; Conservative 0; Mismatches 27;

Qy 1 ccttcaggagcgtcttcggtgcttaagacacccacccttctgtgtatgacatgtgag 60  
||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | |||||  
Db 1007 ccttcgtcttgatgttcttgatcctcagccacagcctccttaataagcattgaggtttgag 1066

Qy 61 cccaaca 67  
||||| |  
Db 1067 ccaaaa 1073

## RESULT 2

US-08-781-891-79/c  
; Sequence 79, Application US/08781891  
; Patent No. 6090620  
; GENERAL INFORMATION:  
; APPLICANT: Fu, Ying-Hui  
; APPLICANT: Yu, Chang-En  
; APPLICANT: Oshima, Junko  
; APPLICANT: Mulligan, John T.  
; APPLICANT: Schellenberg, Gerald D.  
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
; TITLE OF INVENTION: WERNER'S SYNDROME  
; NUMBER OF SEQUENCES: 209  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/781,891  
; FILING DATE: 27-DEC-1996  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 6090620tenburg ph.d., Carol  
; REGISTRATION NUMBER: 39,317  
; REFERENCE/DOCKET NUMBER: 240052.419

TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 79:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 87350 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-08-781-891-79

Query Match 32.2%; Score 23.2; DB 3; Length 87350;  
Best Local Similarity 65.4%; Pred. No. 12;  
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 18 tcggtgcttaagacacccacccttctgtgtatgacatgtgagcccaacagt 69  
||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | |||||  
Db 15725 TCGGTGCTTTGGGAGGCCAAGGGTGGGTATCACCTGAGAGCCCTGGAGT 15674

## RESULT 3

PCT-US95-06815-1  
; Sequence 1, Application PC/TUS9506815  
; GENERAL INFORMATION:  
; APPLICANT:  
; APPLICANT:  
; TITLE OF INVENTION: PURIFIED MYCELIOPHTHORA LACCASES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING SAME  
; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Novo Nordisk of North America, Inc.  
; STREET: 405 Lexington Avenue, Suite 6400  
; COUNTRY: U.S.A.  
; ZIP: 10174-6401  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/06815  
; FILING DATE: 31-May-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/253,781  
; FILING DATE: 03-June-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lowney, Karen A.  
; REGISTRATION NUMBER: 31,274  
; REFERENCE/DOCKET NUMBER: 4184.204-WO

TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 867 0123  
; TELEFAX: 212 867 0298  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3187 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: Myceliophthora thermophila  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: 833...917

FEATURE:  
; NAME/KEY: intron  
; LOCATION: 996...1077

FEATURE:  
; NAME/KEY: intron  
; LOCATION: 1090...1188

FEATURE:  
; NAME/KEY: intron  
; LOCATION: 1261...1332

FEATURE:  
; NAME/KEY: intron  
; LOCATION: 2305...2451

FEATURE:  
; NAME/KEY: intron  
; LOCATION: 2521...2613

FEATURE:  
; NAME/KEY: CDS  
; LOCATION: join (587..832, 918..995, 1078..1089, 1189..1260,  
; LOCATION: 1333..2304, 2452..2520, 2614..3024)  
PCT-US95-06815-1

Query Match 31.9%; Score 23; DB 5; Length 3187;  
Best Local Similarity 63.6%; Pred. No. 5.2;  
Matches 35; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 15 gcttcggtgccttaagacacccacccttctgtgtatgacatgtgagcccaacagt 69  
||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | |||||  
Db 1540 GCGTCGCTGCCGTACGACACCGACCTGGGTGTGTCCCATCAGGACTACTACT 1594

## RESULT 4

US-08-706-037-26  
; Sequence 26, Application US/08706037  
; Patent No. 5770419  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Feng









; OTHER INFORMATION: /note= "HLA-H.CONTIG"  
US-08-724-394A-20

; OTHER INFORMATION: /note= "HLA-H. CONTIG"  
US-08-724-394A-21

|    |                                                                         |                  |               |            |                |
|----|-------------------------------------------------------------------------|------------------|---------------|------------|----------------|
|    | Query Match                                                             | 30.6%            | Score 22;     | DB 2;      | Length 246240; |
|    | Best Local Similarity                                                   | 59.7%;           | Pred. No. 50; |            |                |
|    | Matches                                                                 | 37; Conservative | 0; Mismatches | 25; Indels | 0; Gaps        |
| Qy | gacgatgttcctggcgccctaagaccactaccttttgttgcatacagtgcgcccaaacag            | 68               |               |            |                |
| Dd | 183607 GAAGACCCCTTATGTGTCCTTTCAACTCTCACATCTATTATTAGATCSAATTGGAGTGCCACAG | 183666           |               |            |                |

Search completed: July 31, 2002, 12:21:09  
Job time: 11692 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:11:57 ; Search time 5855.71 Seconds  
(without alignments)  
165.954 Million cell updates/sec

Title: US-09-899-718A-4  
Perfect score: 72  
Sequence: 1 cetttcagacgacgtctgg.....catgtgagcccaacagtggc 72

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estnu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_hic:\*
- 9: gb\_estl:\*
- 10: gb\_est2:\*
- 11: gb\_hic:\*
- 12: gb\_gss:\*
- 13: em\_gss\_hum:\*
- 14: em\_gss\_inv:\*
- 15: em\_gss\_pln:\*
- 16: em\_gss\_vrt:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| C 1        | 27.2  | 37.8        | 450    | 12 | AQ466003 HS-5139_B |
| C 2        | 27.2  | 37.5        | 544    | 10 | BE750138 201331_MA |
| C 3        | 27.2  | 37.5        | 877    | 10 | BF028674 601763910 |
| C 4        | 26.8  | 37.2        | 340    | 12 | BH103249 RPCI-24-2 |
| C 5        | 26.4  | 36.7        | 930    | 10 | BG676499 602622960 |
| C 6        | 26.4  | 36.7        | 2069   | 10 | BG852601 1024034F0 |
| C 7        | 26.3  | 36.1        | 305    | 12 | BH176205 005_F-08  |
| C 8        | 26.3  | 36.1        | 305    | 12 | CNS07J0P           |
| C 9        | 26.3  | 36.1        | 305    | 12 | CNS07J0Q           |
| C 10       | 26.3  | 36.1        | 447    | 12 | AQ536358           |
| C 11       | 26.3  | 36.1        | 557    | 9  | AW836780           |
| C 12       | 26.3  | 36.1        | 622    | 12 | AQ320218           |
| C 13       | 25.8  | 35.8        | 502    | 12 | AQ716839           |
| C 14       | 25.6  | 35.6        | 273    | 9  | BB524309           |
| C 15       | 25.6  | 35.6        | 625    | 10 | BF143122           |
| C 16       | 25.4  | 35.3        | 510    | 12 | AQ785147           |
| C 17       | 25.4  | 35.3        | 555    | 12 | BH210823           |

|      |      |      |      |    |          |
|------|------|------|------|----|----------|
| 18   | 25.4 | 35.3 | 581  | 12 | A2941767 |
| C 19 | 25.2 | 35.0 | 285  | 9  | BA483511 |
| C 20 | 25.2 | 35.0 | 413  | 12 | AQ297755 |
| C 21 | 25.2 | 35.0 | 471  | 12 | AQ005736 |
| C 22 | 25.2 | 35.0 | 482  | 12 | A2230592 |
| C 23 | 25.2 | 34.7 | 127  | 10 | BE854968 |
| C 24 | 25.2 | 34.7 | 193  | 10 | BF183196 |
| C 25 | 25.2 | 34.7 | 235  | 9  | AV336150 |
| C 26 | 25.2 | 34.7 | 253  | 9  | BE140819 |
| C 27 | 25.2 | 34.7 | 270  | 9  | AI592234 |
| C 28 | 25.2 | 34.7 | 394  | 9  | AA169000 |
| C 29 | 25.2 | 34.7 | 465  | 9  | AU179961 |
| C 30 | 25.2 | 34.7 | 518  | 12 | AQ433466 |
| C 31 | 25.2 | 34.7 | 576  | 10 | BM253384 |
| C 32 | 25.2 | 34.7 | 620  | 12 | AQ722509 |
| C 33 | 25.2 | 34.7 | 636  | 10 | BF725277 |
| C 34 | 25.2 | 34.7 | 673  | 9  | BB322109 |
| C 35 | 25.2 | 34.7 | 674  | 10 | BQ20917  |
| C 36 | 25.2 | 34.7 | 747  | 10 | BF664154 |
| C 37 | 25.2 | 34.7 | 769  | 10 | BG484311 |
| C 38 | 25.2 | 34.7 | 1161 | 10 | BG331922 |
| C 39 | 24.8 | 34.4 | 152  | 10 | R56980   |
| C 40 | 24.8 | 34.4 | 231  | 10 | BF955906 |
| C 41 | 24.8 | 34.4 | 406  | 10 | BE524520 |
| C 42 | 24.8 | 34.4 | 635  | 12 | A2111630 |
| C 43 | 24.8 | 34.4 | 642  | 12 | A2951598 |
| C 44 | 24.8 | 34.4 | 661  | 12 | A2496954 |
| C 45 | 24.8 | 34.4 | 663  | 12 | BH211493 |

## ALIGNMENTS

RESULT 1  
AQ466003/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

AQ466003 450 bp DNA linear GSS 23-APR-1999  
HS-5139\_B2\_E07\_SP6E RPCI-11 Human Male BAC Library Homo sapiens  
genomic clone Plate=715 Col=14 Row=J, DNA sequence.  
AQ466003.1 GI:4643098  
GSS.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 450)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)  
or from Resear h Genetics (info@resgen.com). BAC end Web Server:  
http://www.husc.washington.edu  
Plate: 715 row: J column: 14  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 450.  
Location/Qualifiers  
1. .450  
/organism="Homo sapiens"

FEATURES  
source









Email: hbeatig.org  
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs ([info@resgen.com](mailto:info@resgen.com)). BAC end search page: [http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html).  
 Seq primer: SP6  
 Class: BAC ends.

| FEATURES                                                                             | source                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | Location/Qualifiers                                   |
|--------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------|
| 1. .447                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                       |
| /organism="Homo sapiens"                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                       |
| /db_xref="GDB:7622025"                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                       |
| /db_xref="taxon:9606"                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                       |
| /clone="RPCI-11-318M10"                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                       |
| /clone_lib="RPCI-11"                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                       |
| /sex="Male"                                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                       |
| /cell_type="Lymphocytes"                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                       |
| /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library" |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                       |
| BASE COUNT                                                                           | 141 a 91 c 100 g 115 t                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                       |
| ORIGIN                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                       |
| Query Match                                                                          | 36.1%                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Score 26; DB 12; Length 447;                          |
| Best Local Similarity                                                                | 70.0%                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Prod. No. 70;                                         |
| Matches                                                                              | 35; Conservative                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 0; Mismatches 15; Indels 0; Gaps 0;                   |
| QY                                                                                   | 16                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | cttcgggtccttaagacacctacctcttgcctatgcacatgtgagcccaa 65 |
|                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                       |
| DB                                                                                   | 310                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | CTTTCATCATCATAGTCATCATAGTGTGTGTCTCACATGTGAGGTCAA 359  |
| RESULT                                                                               | 11                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                       |
| AW836780                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | 557 bp mRNA linear EST 18-MAY-2000                    |
| LOCUS                                                                                | QV1-LH0036-150200-070-f01                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | LH0036 Homo sapiens cDNA, mRNA sequence.              |
| DEFINITION                                                                           | AW836780                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                       |
| ACCESSION                                                                            | AW836780.1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | GI:7930754                                            |
| VERSION                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                       |
| KEYWORDS                                                                             | EST.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                       |
| SOURCE                                                                               | human.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                       |
| ORGANISM                                                                             | Homo sapiens                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                       |
|                                                                                      | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                       |
| REFERENCE                                                                            | 1 (bases 1 to 557)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                       |
| AUTHORS                                                                              | Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.                                                                                                                                                                                                                                                                                                                                                                                                                           |                                                       |
| TITLE                                                                                | Shotgun sequencing of the human transcriptome with ORF expressed sequence tags                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                       |
| JOURNAL                                                                              | Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                       |
| MEDLINE                                                                              | 20202663                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                       |
| COMMENT                                                                              | Contact: Simpson A.J.G.<br>Laboratory of Cancer Genetics<br>Ludwig Institute for Cancer Research<br>Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil<br>Tel: +55-11-2704922<br>Fax: +55-11-2707001<br>Email: <a href="mailto:asimpson@ludwig.org.br">asimpson@ludwig.org.br</a><br>This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL ( <a href="http://www.ludwig.org.br/scripts/gethtml2.pl?tbl=st2-QV1-LH0036-150200-070-f01&amp;t3=2000-02-15&amp;t4=1">http://www.ludwig.org.br/scripts/gethtml2.pl?tbl=st2-QV1-LH0036-150200-070-f01&amp;t3=2000-02-15&amp;t4=1</a> )<br>Seq primer: puc 18 forward<br>High quality sequence start: 11<br>High quality sequence stop: 557. |                                                       |
| FEATURES                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                       |
| source                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Location/Qualifiers                                   |
|                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | 1. .557                                               |

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="L70036"
/dev_stage="Adult"
/note="Organ: leiomyos; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORSTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      172 a 126 c 118 g 141 t
ORIGIN

```

```

Query Match      36.1%; Score 26; DB 9; Length 557;
Best Local Similarity 70.0%; Pred. No. 75;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```

```

QY 16 ctctggtgccttaagacacacctctgtgtctatgacatgtgagccca 65
      ||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 272 CTTTCATCACATAAGTCATCTACAGTTGTGTCTCAGTGTGAGTCAA 321

```

## RESULT 12

```

AQ320218      AQ320218      622 bp      DNA      linear      GSS 06-MAY-1999
LOCUS      RPC111-100C1.TV RPC1-11 Homo sapiens genomic clone RPC1-11-100C1,
DEFINITION      DNA sequence.
ACCESSION      AQ320218
VERSION      AQ320218.1 GI:4053029
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens

```

```

REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
      Adams M.D., Rounsley S.D., Zhao S., Bass S., Linher, K., Golden, K.,
      Barry, K., Granger D., Suh, E., Wible, C., de Jong, P. and Venter, J.C.
      Use of human BAC End Sequences for Sequence-Ready Map Building
      Unpublished (1998)
      Other_GSSs: RPC111-100C1.TJ
      Contact: Shaying Zhao, William Nierman, Mark Adams
      Department of Eukaryotic Genomics
      The Institute for Genomic Research
      9712 Medical Center Dr., Rockville, MD 20850
      Tel: 301 838 0200
      Fax: 301 838 0208
      Email: hbe@tigr.org

```

```

Clones are derived from the human BAC library RPC1-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.
FEATURES
      source
      Location/Qualifiers
      1..622
      /organism="Homo sapiens"
      /db_xref="GDB:7538064"
      /db_xref="taxon:9606"
      /clone="RPC1-11-100C1"
      /clone_lib="RPC1-11"
      /sex="Male"
      /cell_type="Lymphocytes"
      /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
      RPC11 Human Male BAC Library"
BASE COUNT      195 a 129 c 134 g 164 t
ORIGIN

```

```

Query Match      36.1%; Score 26; DB 12; Length 622;

```

```

Best Local Similarity 70.0%; Pred. No. 77;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```

```

QY 16 ctctggtgccttaagacacacctctgtgtctatgacatgtgagccca 65
      ||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 304 CTTTCATCACATAAGTCATCTACAGTTGTGTCTCAGTGTGAGTCAA 353

```

## RESULT 13

```

AQ716839      AQ716839      502 bp      DNA      linear      GSS 13-JUL-1999
LOCUS      HS_5473_AL_D10_17A RPC1-11 Human Male BAC library Homo sapiens
DEFINITION      genomic clone Plate=1049 Col=19 Row=G, DNA sequence.
ACCESSION      AQ716839
VERSION      AQ716839.1 GI:5466155
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens

```

```

REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
      1 (bases 1 to 502)
      Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
      Kellier, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
      Hood, L.
      Sequence-tagged connectors: A sequence approach to mapping and
      scanning the human genome
      Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
      99380589
      Contact: Mahairas GG, Wallace JC, Hood L
      High Throughput Sequencing Center
      University of Washington
      401 Queen Anne Avenue North, Seattle, WA 98109, USA
      Tel: (206) 616-3618
      Fax: (206) 616-3887
      Email: jwallace@u.washington.edu

```

```

Clones are derived from the human BAC library RPC1-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research h Genetics (info@resgen.com). BAC end Web Server:
http://www.hsc.washington.edu
Plate: 1049 row: G column: 19
Seq primer: T7
Class: BAC ends
High quality sequence stop: 502.
FEATURES
      source
      Location/Qualifiers
      1..502
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="Plate=1049 Col=19 Row=G"
      /clone_lib="RPC1-11 Human Male BAC Library"
      /sex="male"
      /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
      Male blood DNA was isolated from one randomly chosen donor
      and partially digested with a combination of EcoRI and
      EcoRI Methylase. Size selected DNA was cloned into the
      pBACE3.6 vector at EcoRI sites"
BASE COUNT      130 a 133 c 100 g 134 t
ORIGIN

```

```

Query Match      35.8%; Score 25.8; DB 12; Length 502;
Best Local Similarity 67.9%; Pred. No. 85;
Matches 36; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```

```

QY 2 ctctcagacacatcttcggtgccttaagacacacctctgtgtctatgaca 54
      | | | | | | | | | | | | | | | | | | | | | | | | |
Db 317 CTTCTGGGATATGACTGTGTTCCCAAGGCGCCTCTTTGGGTCTAAGACA 369

```

## RESULT 14

```

BB524309/c      BB524309      273 bp      mRNA      linear      EST 28-JUL-2000
LOCUS

```

Thu Aug 1 08:30:01 2002

|            |                                                                                                                |
|------------|----------------------------------------------------------------------------------------------------------------|
| DEFINITION | BB524309 RIKEN full-length enriched, 15 days embryo head Mus musculus cDNA clone D930019D20 3', mRNA sequence. |
| ACCESSION  | BB524309                                                                                                       |
| VERSION    | BB524309.1                                                                                                     |
| KEYWORDS   | EST<br>GI:9575767                                                                                              |
| BASE COUNT | 64 a 75 c 63 g 71 t                                                                                            |
| ORIGIN     | FLC I"                                                                                                         |
|            | was cloned into the XhoI and BamHI sites. Vector: a modified pbluescript KS(+) after bulk excision from Lambda |

| SOURCE       | ORGANISM                                                           | Query Match           | Score  | DB           | Length |
|--------------|--------------------------------------------------------------------|-----------------------|--------|--------------|--------|
| house mouse. | Mus musculus                                                       | 35.6%                 | 25.6;  | 9;           | 273;   |
|              | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;            |                       |        |              |        |
|              | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. |                       |        |              |        |
|              | 1 (bases 1 to 273)                                                 | Best Local Similarity | 59.7%; |              |        |
|              |                                                                    | Matches               | 43;    | Conservative | 0;     |
|              |                                                                    |                       |        | Mismatches   | 29;    |
|              |                                                                    |                       |        | Indels       | 0;     |
|              |                                                                    |                       |        | Gaps         | 0;     |

[illegible]

|         |                                                                                                                                                                                                                                                              |                                                                                    |                 |
|---------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|-----------------|
| TITLE   | RIKEN Mouse ESTs (Konno, H., et al.)                                                                                                                                                                                                                         |                                                                                    |                 |
| JOURNAL | Unpublished (2000)                                                                                                                                                                                                                                           |                                                                                    |                 |
| COMMENT | Contact: Yoshihide Havashizaki                                                                                                                                                                                                                               |                                                                                    |                 |
|         | Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tominaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamana, I., Yano, K., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y. |                                                                                    |                 |
|         | LOCUS                                                                                                                                                                                                                                                        | BF143122                                                                           | 625 bp          |
|         | DEFINITION                                                                                                                                                                                                                                                   | 601788012F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4015898 5', mRNA sequence. | linear mRNA     |
|         | RESULT                                                                                                                                                                                                                                                       | 15                                                                                 | EST 24-OCT-2000 |
|         | BF143122                                                                                                                                                                                                                                                     |                                                                                    |                 |

|                                                                                                                                                                                                                                                                                                                                  |                                                        |                                                                |             |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------|----------------------------------------------------------------|-------------|
| Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute<br>The Institute of Physical and Chemical Research (RIKEN)<br>1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan<br>Tel.: 81-45-503-9222<br>Fax: 81-45-503-9216<br>Email: genome-res@gs.c.riken.go.jp, | ACCESSION<br>VERSION<br>KEYWORDS<br>SOURCE<br>ORGANISM | BF143122<br>BF143122.1<br>EST.<br>house mouse.<br>Mus musculus | GI:10982162 |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------|----------------------------------------------------------------|-------------|

|                                                                                                                                                                                              |                                                                                                                                                              |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| URL: <a href="http://genome.qsc.riken.go.jp/">http://genome.qsc.riken.go.jp/</a>                                                                                                             | REFERENCE 1 (bases 1 to 625)                                                                                                                                 |
| Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.                                                                 | AUTHORS NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .                                                                              |
| Thermotabolization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998) | TITLE National Institutes of Health, Mammalian Gene Collection (MGC)                                                                                         |
| Itoh, M., Kitsuunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,                                                                                                                     | JOURNAL Unpublished (1999)                                                                                                                                   |
|                                                                                                                                                                                              | COMMENT Contact: Robert Strausberg, Ph. D. Email: <a href="mailto:cgapbs-r@mail.nih.gov">cgapbs-r@mail.nih.gov</a> Tissue Procurement: Gilbert Smith, Ph. D. |

Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation system. *Genome Res.* 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303, 19-44 (1999)

|                     |                          |
|---------------------|--------------------------|
| FEATURES            |                          |
| source              | 1..273                   |
| Location/Qualifiers | /organism="Mus musculus" |
|                     | /strain="C57BL/6J"       |
|                     | /db_xref="taxon:10090"   |
|                     | /db_xref="taxon:10090"   |
| FEATURES            |                          |
| source              | 1..625                   |
| Location/Qualifiers | /organism="Mus musculus" |
|                     | /strain="CZECH II"       |
|                     | /db_xref="taxon:10090"   |
|                     | /clone="IMAGE:4018698"   |
| FEATURES            |                          |
| source              | 1..273                   |
| Location/Qualifiers | /organism="Mus musculus" |
|                     | /strain="C57BL/6J"       |
|                     | /db_xref="taxon:10090"   |
|                     | /db_xref="taxon:10090"   |

```

/clone="D930019D20"
/clone_lib="RIKEN full-length enriched, 15 days embryo
head"
/sex="mixed"
/tissue_type="head"
/dev_stage="15 days embryo"
/lab_host="DH10B"

/clone="NCI_CGAP_Lu30"
/tissue_type="tumor, metastatic to mammary"
/lab_host="DH10B"
/notes="Organ lung; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; transgenic model WNT-1, expression driven by
MTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
dr. Library constructed by Life Technologies. WNT1"

```

```

/note=*Site_1: Sali; Site2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',

```

GAGAGAGAAGATCTCGAGTTAATAATTCGCCCCCCCCC 3'}. cdna  
GAGAGAGAAGATCTCGAGTTAATAATTCGCCCCCCCCC 3'}. cdna  
primer adaptor of sequence [5',  
cap-trapper. Second strand cDNA was prepared with the  
transcriptase and subsequently enriched for full-length by  
prepared by using trehalose thermo-activated reverse  
GAGAGAGAAGATCTCGAGTTAATAATTCGCCCCCCCCC 3'}. cdna was

|    | Matches | 34;                                              | Conservative | Mismatches | 0; | Indels | 14; | Gaps | 0; |
|----|---------|--------------------------------------------------|--------------|------------|----|--------|-----|------|----|
| Oy | 1       | ccttcaggacgatgcttcgggtccttaagcacaccactctt        | gtgtct       | 48         |    |        |     |      |    |
|    |         |                                                  |              |            |    |        |     |      |    |
| Db | 578     | CCTTTTACCACAATCCTCTGTAACTTAAGAAACCTTGAAATTATGTCT | 625          |            |    |        |     |      |    |

10

Thu Aug 1 08:30:01 2002

us-09-899-718a-4.rst

Page 8

Search completed: July 31, 2002, 12:12:01  
Job time: 13474 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:11:11 ; Search time 720.8 Seconds  
(without alignments)  
171.501 Million cell updates/sec

Title: us-09-899-718a-4  
Perfect score: 72  
Sequence: 1 ctttcagagcagctctg.....catgtgagcccaacagtgc 72

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No.              | Score | Query Match | Length | ID | Description |
|-------------------------|-------|-------------|--------|----|-------------|
| 1                       | 25.6  | 35.6        | 3085   | 23 | ABL16797    |
| 2                       | 25.6  | 35.6        | 20261  | 23 | ABL16796    |
| 3                       | 25    | 34.7        | 152    | 21 | AAC07947    |
| 4                       | 25    | 34.7        | 424    | 21 | AAC00215    |
| 5                       | 25    | 34.7        | 626    | 20 | AA13569     |
| 6                       | 25    | 34.7        | 907    | 20 | AA253005    |
| 7                       | 25    | 34.7        | 3098   | 22 | ABA18125    |
| 8                       | 25    | 34.7        | 3098   | 22 | AA540421    |
| 9                       | 25    | 34.7        | 3098   | 22 | AA104024    |
| Drosophila melanogaster |       |             |        |    |             |
| Drosophila melanogaster |       |             |        |    |             |
| Human secreted pro      |       |             |        |    |             |
| Enterococcus faeca      |       |             |        |    |             |
| Human prostate tum      |       |             |        |    |             |
| Human nervous syst      |       |             |        |    |             |
| DNA encoding human      |       |             |        |    |             |
| Human reproductive      |       |             |        |    |             |

|      |      |      |        |    |           |                    |
|------|------|------|--------|----|-----------|--------------------|
| c 10 | 25   | 34.7 | 3098   | 22 | AAK86871  | Human immune/haema |
| c 11 | 25   | 34.7 | 4316   | 22 | ABA83122  | H1A-Cw ovarian tum |
| c 12 | 24.6 | 34.2 | 50000  | 24 | AAAD26400 | Human glutamate re |
| c 13 | 24.6 | 34.2 | 50000  | 24 | AAAD26437 | Human GRM3 gene fr |
| c 14 | 24.2 | 33.6 | 479    | 22 | ABA43910  | Human breast cell  |
| c 15 | 24.2 | 33.6 | 479    | 22 | ABA54363  | Human foetal liver |
| c 16 | 24.2 | 33.6 | 479    | 22 | ABA24139  | Probe #2605 for ge |
| c 17 | 24.2 | 33.6 | 479    | 22 | AAK02657  | Human brain expres |
| c 18 | 24.2 | 33.6 | 479    | 22 | AAK28093  | Human bone marrow  |
| c 19 | 24.2 | 33.6 | 479    | 22 | AAI12672  | Probe #2605 for ge |
| c 20 | 24.2 | 33.6 | 479    | 22 | AAI34021  | Probe #2707 used t |
| c 21 | 24.2 | 33.6 | 479    | 22 | AAI02580  | Probe #2571 used t |
| c 22 | 24.2 | 33.6 | 15464  | 23 | ABL06682  | Drosophila melanog |
| c 23 | 24   | 33.3 | 388    | 22 | AAI86498  | Human polynucleoti |
| c 24 | 23.8 | 33.1 | 2529   | 17 | AAI31531  | Human 3' apolipop  |
| c 25 | 23.8 | 33.1 | 2529   | 21 | AAI31997  | Human apolipoprote |
| c 26 | 23.4 | 32.5 | 273    | 22 | AAI29187  | Colon tumour relat |
| c 27 | 23.4 | 32.5 | 409    | 23 | AAI39790  | DNA encoding novel |
| c 28 | 23.4 | 32.5 | 435    | 22 | AAI29916  | Human lung antigen |
| c 29 | 23.4 | 32.5 | 747    | 23 | AAI58532  | cDNA #1208 encodin |
| c 30 | 23.4 | 32.5 | 2016   | 11 | AAQ02887  | cDNA of human plac |
| c 31 | 23.4 | 32.5 | 2360   | 11 | AAQ02888  | Human placenta-der |
| c 32 | 23.4 | 32.5 | 3300   | 21 | AAO5446   | Streptococcus pneu |
| c 33 | 23.4 | 32.5 | 6588   | 22 | AAH62768  | Shrimp white spot  |
| c 34 | 23.4 | 32.5 | 15213  | 19 | AAV52159  | Streptococcus pneu |
| c 35 | 23.4 | 32.5 | 305107 | 22 | AAH62689  | Shrimp white spot  |
| c 36 | 23.2 | 32.2 | 11628  | 20 | AAI20520  | Polynucleotide seq |
| c 37 | 23.2 | 32.2 | 87350  | 18 | AAI83003  | Human WRN genomic  |
| c 38 | 23   | 31.9 | 3012   | 22 | AAH98552  | Human ESR-derived  |
| c 39 | 23   | 31.9 | 3187   | 17 | AAI10922  | Laccase gene. Myc  |
| c 40 | 23   | 31.9 | 3192   | 18 | AAI63318  | Myceliophthora the |
| c 41 | 23   | 31.9 | 3192   | 21 | AAZ24236  | M. thermophila lac |
| c 42 | 23   | 31.9 | 6868   | 22 | AAI16293  | Human App-murine n |
| c 43 | 23   | 31.9 | 11036  | 20 | AAI3401   | Enterococcus faeca |
| c 44 | 23   | 31.9 | 57728  | 22 | AAI87588  | Human 9p11 chromos |
| c 45 | 22.8 | 31.7 | 1001   | 21 | AAH51258  | Human GSHS related |

## ALIGNMENTS

### RESULT 1

|    |          |                                                                         |
|----|----------|-------------------------------------------------------------------------|
| ID | ABL16797 | standard; DNA; 3085 BP.                                                 |
| XX | AC       | ABL16797;                                                               |
| XX | AC       | ABL16797;                                                               |
| XX | DT       | 26-MAR-2002 (first entry)                                               |
| XX | XX       | Drosophila melanogaster genomic polynucleotide SEQ ID NO 1864.          |
| XX | DE       | Drosophila; developmental biology; cell signalling; insecticide;        |
| XX | KW       | pharmaceutical; gene; ds.                                               |
| XX | KW       | Drosophila melanogaster.                                                |
| OS | XX       | WO200171042-A2.                                                         |
| PN | XX       | 27-SEP-2001.                                                            |
| PD | XX       | 23-MAR-2001; 2001WO-US09231.                                            |
| PF | XX       | 23-MAR-2000; 2000US-191637P.                                            |
| PR | XX       | 11-JUL-2000; 2000US-0614150.                                            |
| XX | XX       | (PEKE ) PE CORP NY.                                                     |
| PA | XX       | Venter JC, Adams M, Li PWD, Myers EW;                                   |
| XX | XX       | WPI; 2001-656860/75.                                                    |
| DR | XX       | New isolated nucleic acid detection reagent for detecting 1000 or more  |
| PT | XX       | genes from Drosophila and for elucidating cell signalling and cell-cell |





|                       |                                                                         |               |    |
|-----------------------|-------------------------------------------------------------------------|---------------|----|
| DT                    | 14-MAR-2000                                                             | (first entry) | XX |
| DE                    | Human prostate tumor cDNA library derived EST fragment #148.            |               | PR |
| DE                    | Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;        |               | PR |
| KW                    | treatment; ds.                                                          |               | PR |
| KW                    |                                                                         |               | PR |
| KW                    |                                                                         |               | PR |
| OS                    | Homo sapiens.                                                           |               | PR |
| OS                    |                                                                         |               | PR |
| XX                    | DE19820190-A1.                                                          |               | PR |
| XX                    |                                                                         |               | PR |
| XX                    | 04-NOV-1999.                                                            |               | PR |
| PD                    |                                                                         |               | PR |
| PD                    | 28-APR-1998; 98DE-1020190.                                              |               | PR |
| PF                    |                                                                         |               | PR |
| XX                    | 28-APR-1998; 98DE-1020190.                                              |               | PR |
| XX                    | (META-) METAGEN GES GENOMFORSCHUNG MBH.                                 |               | PR |
| PA                    |                                                                         |               | PR |
| PI                    | Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;       |               | PR |
| PI                    | WPI; 1999-621386/54.                                                    |               | PR |
| DR                    |                                                                         |               | PR |
| XX                    | New human nucleic acid sequences from pancreatic tumors, and related    |               | PR |
| PT                    | proteins -                                                              |               | PR |
| PT                    |                                                                         |               | PR |
| PS                    | Claim 2; Page 299-300; 502pp; German.                                   |               | PR |
| PS                    |                                                                         |               | PR |
| XX                    |                                                                         |               | PR |
| CC                    | This invention describes novel polypeptides and their encoding nucleic  |               | PR |
| CC                    | acids derived from human pancreatic tumor tissue which have cytostatic  |               | PR |
| CC                    | activity. The sequences are also useful in producing pharmaceutical     |               | PR |
| CC                    | compositions for treatment of pancreatic tumors. AA52858-Z53014         |               | PR |
| CC                    | represent expressed sequence tag (EST) fragments derived from a human   |               | PR |
| CC                    | pancreatic tumor cDNA library and which encode the proteins represented |               | PR |
| CC                    | in AAY73814-174252.                                                     |               | PR |
| XX                    |                                                                         |               | PR |
| XX                    | Sequence 907 BP; 152 A; 266 C; 323 G; 166 T; 0 other;                   |               | PR |
| SQ                    |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               | PR |
| Matches               |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               | PR |
| Matches               |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               | PR |
| Matches               |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               | PR |
| Matches               |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               | PR |
| Matches               |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               | PR |
| Matches               |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               | PR |
| Matches               |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               | PR |
| Matches               |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               | PR |
| Matches               |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               | PR |
| Matches               |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               | PR |
| Matches               |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               | PR |
| Matches               |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               | PR |
| Matches               |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               | PR |
| Matches               |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               | PR |
| Matches               |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               | PR |
| Matches               |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               | PR |
| Matches               |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               | PR |
| Matches               |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               | PR |
| Matches               |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               | PR |
| Matches               |                                                                         |               | PR |
| Query Match           |                                                                         |               | PR |
| Best Local Similarity |                                                                         |               |    |





|    |              |                 |
|----|--------------|-----------------|
| PR | 26-JUL-2000; | 2000US-0220963. |
| PR | 26-JUL-2000; | 2000US-0220964. |
| PR | 14-AUG-2000; | 2000US-0220958. |
| PR | 14-AUG-2000; | 2000US-0224518. |
| PR | 14-AUG-2000; | 2000US-0224519. |
| PR | 14-AUG-2000; | 2000US-0225213. |
| PR | 14-AUG-2000; | 2000US-0225214. |
| PR | 14-AUG-2000; | 2000US-0225216. |
| PR | 14-AUG-2000; | 2000US-0225267. |
| PR | 14-AUG-2000; | 2000US-0225268. |
| PR | 14-AUG-2000; | 2000US-0225270. |
| PR | 14-AUG-2000; | 2000US-0225447. |
| PR | 14-AUG-2000; | 2000US-0225575. |
| PR | 14-AUG-2000; | 2000US-0225758. |
| PR | 14-AUG-2000; | 2000US-0225759. |
| PR | 18-AUG-2000; | 2000US-0226279. |
| PR | 22-AUG-2000; | 2000US-0226681. |
| PR | 22-AUG-2000; | 2000US-0226682. |
| PR | 22-AUG-2000; | 2000US-0226712. |
| PR | 23-AUG-2000; | 2000US-0227009. |
| PR | 23-AUG-2000; | 2000US-0227192. |
| PR | 01-SEP-2000; | 2000US-0228294. |
| PR | 01-SEP-2000; | 2000US-0229287. |
| PR | 01-SEP-2000; | 2000US-0229343. |
| PR | 01-SEP-2000; | 2000US-0229344. |
| PR | 05-SEP-2000; | 2000US-0229345. |
| PR | 05-SEP-2000; | 2000US-0229513. |
| PR | 06-SEP-2000; | 2000US-0229513. |
| PR | 06-SEP-2000; | 2000US-0230437. |
| PR | 08-SEP-2000; | 2000US-0230438. |
| PR | 08-SEP-2000; | 2000US-0231242. |
| PR | 08-SEP-2000; | 2000US-0231243. |
| PR | 08-SEP-2000; | 2000US-0231244. |
| PR | 08-SEP-2000; | 2000US-0231413. |
| PR | 08-SEP-2000; | 2000US-0231414. |
| PR | 08-SEP-2000; | 2000US-0232080. |
| PR | 08-SEP-2000; | 2000US-0232081. |
| PR | 12-SEP-2000; | 2000US-0231968. |
| PR | 14-SEP-2000; | 2000US-0232397. |
| PR | 14-SEP-2000; | 2000US-0232398. |
| PR | 14-SEP-2000; | 2000US-0232399. |
| PR | 14-SEP-2000; | 2000US-0232400. |
| PR | 14-SEP-2000; | 2000US-0232401. |
| PR | 14-SEP-2000; | 2000US-0232063. |
| PR | 14-SEP-2000; | 2000US-0233064. |
| PR | 21-SEP-2000; | 2000US-0233065. |
| PR | 21-SEP-2000; | 2000US-0234223. |
| PR | 21-SEP-2000; | 2000US-0234274. |
| PR | 25-SEP-2000; | 2000US-0234297. |
| PR | 25-SEP-2000; | 2000US-0234998. |
| PR | 26-SEP-2000; | 2000US-0235484. |
| PR | 27-SEP-2000; | 2000US-0235834. |
| PR | 27-SEP-2000; | 2000US-0235836. |
| PR | 27-SEP-2000; | 2000US-0236327. |
| PR | 29-SEP-2000; | 2000US-0236367. |
| PR | 29-SEP-2000; | 2000US-0236368. |
| PR | 29-SEP-2000; | 2000US-0236369. |
| PR | 29-SEP-2000; | 2000US-0236370. |
| PR | 02-OCT-2000; | 2000US-0236802. |
| PR | 02-OCT-2000; | 2000US-0237038. |
| PR | 02-OCT-2000; | 2000US-0237037. |
| PR | 02-OCT-2000; | 2000US-0237039. |
| PR | 02-OCT-2000; | 2000US-0237040. |
| PR | 13-OCT-2000; | 2000US-0239935. |
| PR | 13-OCT-2000; | 2000US-0239937. |
| PR | 20-OCT-2000; | 2000US-0240960. |
| PR | 20-OCT-2000; | 2000US-0241121. |
| PR | 20-OCT-2000; | 2000US-0241785. |
| PR | 20-OCT-2000; | 2000US-0241786. |
| PR | 20-OCT-2000; | 2000US-0241787. |
| PR | 20-OCT-2000; | 2000US-0241808. |
| PR | 20-OCT-2000; | 2000US-0241809. |
| PR | 20-OCT-2000; | 2000US-0241826. |
| PR | 01-NOV-2000; | 2000US-0244617. |
| PR | 08-NOV-2000; | 2000US-0244674. |

|    |              |                         |  |
|----|--------------|-------------------------|--|
| PR | 08-NOV-2000; | 2000US-0246472;         |  |
| PR | 08-NOV-2000; | 2000US-0246476;         |  |
| PR | 08-NOV-2000; | 2000US-0246477;         |  |
| PR | 08-NOV-2000; | 2000US-0246478;         |  |
| PR | 08-NOV-2000; | 2000US-0246523;         |  |
| PR | 08-NOV-2000; | 2000US-0246524;         |  |
| PR | 08-NOV-2000; | 2000US-0246525;         |  |
| PR | 08-NOV-2000; | 2000US-0246526;         |  |
| PR | 08-NOV-2000; | 2000US-0246527;         |  |
| PR | 08-NOV-2000; | 2000US-0246528;         |  |
| PR | 08-NOV-2000; | 2000US-0246532;         |  |
| PR | 08-NOV-2000; | 2000US-0246609;         |  |
| PR | 08-NOV-2000; | 2000US-0246611;         |  |
| PR | 08-NOV-2000; | 2000US-0246613;         |  |
| PR | 17-NOV-2000; | 2000US-0249207;         |  |
| PR | 17-NOV-2000; | 2000US-0249208;         |  |
| PR | 17-NOV-2000; | 2000US-0249209;         |  |
| PR | 17-NOV-2000; | 2000US-0249210;         |  |
| PR | 17-NOV-2000; | 2000US-0249211;         |  |
| PR | 17-NOV-2000; | 2000US-0249212;         |  |
| PR | 17-NOV-2000; | 2000US-0249213;         |  |
| PR | 17-NOV-2000; | 2000US-0249214;         |  |
| PR | 17-NOV-2000; | 2000US-0249216;         |  |
| PR | 17-NOV-2000; | 2000US-0249217;         |  |
| PR | 17-NOV-2000; | 2000US-0249218;         |  |
| PR | 17-NOV-2000; | 2000US-0249244;         |  |
| PR | 17-NOV-2000; | 2000US-0249245;         |  |
| PR | 17-NOV-2000; | 2000US-0249266;         |  |
| PR | 17-NOV-2000; | 2000US-0249297;         |  |
| PR | 17-NOV-2000; | 2000US-0249299;         |  |
| PR | 17-NOV-2000; | 2000US-0249300;         |  |
| PR | 01-DEC-2000; | 2000US-0250160;         |  |
| PR | 01-DEC-2000; | 2000US-0250391;         |  |
| PR | 05-DEC-2000; | 2000US-0251030;         |  |
| PR | 05-DEC-2000; | 2000US-0251988;         |  |
| PR | 05-DEC-2000; | 2000US-0256719;         |  |
| PR | 06-DEC-2000; | 2000US-0251479;         |  |
| PR | 08-DEC-2000; | 2000US-0251856;         |  |
| PR | 08-DEC-2000; | 2000US-0251868;         |  |
| PR | 08-DEC-2000; | 2000US-0251869;         |  |
| PR | 08-DEC-2000; | 2000US-0251989;         |  |
| PR | 08-DEC-2000; | 2000US-0251990;         |  |
| PR | 11-DEC-2000; | 2000US-0254097;         |  |
| PR | 05-JAN-2001; | 2001US-0259678;         |  |
| XX | (HUMAN - )   | HUMAN GENOME PROJ. INC. |  |

(HUMA- ) HUMAN GENOME SCI. INC.

Rosen CA. Barash SC. Ruben SM:

WPT: 2001-451929/48.

Isolated polypeptide for treating, preventing and/or prognosing disorders related to the reproductive system including prostate cancer and also for testing and detection e.g. diagnosis -

Disclosure: SEO ID No 573: 546pp: English.

The invention relates to novel isolated human prostate cancer antigen polynucleotides (I) and polypeptides (II). (I) and (II) are useful for preventing, treating or ameliorating a medical condition when administered. (I), (II) and the antibody to (II) are useful for treating, preventing and/or prognosing disorders related to the reproductive system including prostate cancers; urinary disorders e.g. chronic nephritis; and blood-related disorders e.g. thrombosis. (II) can be used for testing and detection e.g. as a chromosomal marker and in forensics. (I) and the anti-(II) antibody can be used in testing and detection in immunoassays. AAS40061-AAS40775 represent the human prostate cancer antigen coding sequences, and related PCR primers and sequences of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format



```

PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition
XX
XX Disclosure; SEQ ID NO 6712; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
XX Sequence 3098 BP; 629 A; 858 C; 936 G; 675 T; 0 other;
SQ
Query Match 34.7%; Score 25; DB 22; Length 3098;
Best Local Similarity 69.4%; Pred No. 6.4;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 6 caggacgatgcttcggtgacctgaacacacctcttggctgtatgaca 54
    ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 57 CAGGAGGAGGCTTCGGGGGCCCATGACCCCAACCTCCGCTCTGGGAAA 9
RESULT 10
AAK86871/c
ID AAK86871 standard; DNA; 3098 BP.
XX
XX AAK86871;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:41683.
DE
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
PD
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218390.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.

```





AC AAD26437;  
 XX 26-MAR-2002 (first entry)  
 XX Human GRM3 gene fragment #1 allelic variant containing polymorphic site.  
 XX Human; glutamate receptor metabotropic 3; GRM3; neuroprotective;  
 XX nootropic; gene therapy; neurotransmitter; chromosome 7q21.1-q21.2;  
 XX drug screening; neurological disorder; polymorphic site; PS; ds.  
 XX Homo sapiens.  
 XX Key Location/Qualifiers  
 XX misc\_feature 4006  
 XX /\*tag= a  
 XX /note= "This degenerate base represents polymorphic site  
 XX (PS) 1"  
 XX 4192  
 XX misc\_feature  
 XX /\*tag= b  
 XX /note= "This degenerate base represents polymorphic site  
 XX (PS) 2"  
 XX 25586  
 XX misc\_feature  
 XX /\*tag= c  
 XX /note= "This degenerate base represents polymorphic site  
 XX (PS) 3"  
 XX 26137  
 XX misc\_feature  
 XX /\*tag= d  
 XX /note= "This degenerate base represents polymorphic site  
 XX (PS) 4"  
 XX WO200196350-A2.  
 XX 20-DEC-2001.  
 XX 18-JUN-2001; 2001WO-US19447.  
 XX 16-JUN-2000; 2000US-212328P.  
 XX (GENA-) GENAISSANCE PHARM INC.  
 XX Finkel K, Koshy B, Tanguay DA;  
 XX WPI; 2002-090198/12.  
 XX New isolated polynucleotide, a polymorphic variant of glutamate  
 XX receptor, metabotropic 3 (GRM3) gene for expressing GRM3 protein  
 XX isoform to screen drugs to treat GRM3 activity-related disease -  
 XX Claim 20; Page 135-153; 165pp; English.  
 XX The invention relates to an isolated polynucleotide which is a  
 XX polymorphic variant of glutamate receptor, metabotropic 3 (GRM3)  
 XX isogene. GRM3 is a receptor for glutamate, the major excitatory  
 XX neurotransmitter in the mammalian central nervous system. Human GRM3  
 XX located on chromosome 7q21.1-q21.2 is expressed in human foetal and adult  
 XX whole brain especially in the caudate nucleus and corpus collosum. GRM3  
 XX DNA is useful in gene therapy and also for studying the expression and  
 XX function of GRM3. GRM3 polypeptide is used for screening drugs. A  
 XX recombinant non-human organism is used to study expression of GRM3 SG  
 XX in vivo, for in vivo screening and testing of drugs targetted against  
 XX GRM3 protein, and for testing the efficacy of therapeutic agents and  
 XX compounds for neurological disorders in a biological system. GRM3  
 XX haplotypes are for treating diseases associated with GRM3 activity.  
 XX e.g., neurological disorders. The present sequence is an allelic variant  
 XX of human GRM3 gene fragment containing polymorphic sites (PS).  
 XX Sequence 50000 BP; 15337 A; 9554 C; 9287 G; 15818 T; 4 other;

Query Match 34.2%; Score 24.6; DB 24; Length 50000;  
 Best Local Similarity 65.5%; Pred. No. 22;  
 Matches 36; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Qy 7 agagcagatgcttcgggtccttaagacacaccccttctgtctgtcatgtgagc 61  
 ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 27892 AGGAACATGTTTGGTTCCTAARACACCTTCGAAGAGCTCTCTGATTTCTAAGC 27838  
 RESULT 14  
 ABA43910/C  
 ID ABA43910 standard; DNA; 479 BP.  
 XX ABA43910;  
 XX 01-FEB-2002 (first entry)  
 XX Human breast cell single exon nucleic acid probe #2605.  
 XX Human; microarray; single exon probe; gene expression; breast;  
 XX disease; cancer; ss.  
 XX Homo sapiens.  
 XX WO200157271-A2.  
 XX 09-AUG-2001.  
 XX 30-JAN-2001; 2001WO-US00662.  
 XX 04-FEB-2000; 2000US-0180312.  
 XX 26-MAY-2000; 2000US-0207456.  
 XX 30-JUN-2000; 2000US-0608408.  
 XX 03-AUG-2000; 2000US-0632366.  
 XX 21-SEP-2000; 2000US-0234687.  
 XX 27-SEP-2000; 2000US-0236359.  
 XX 04-OCT-2000; 2000GB-0024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-496933/54.  
 XX New spatially-addressable set of single exon nucleic acid probes,  
 XX useful for measuring gene expression in sample derived from human  
 XX breast, comprises number of single exon nucleic acid probes -  
 XX Claim 1; SEQ ID NO 2605; 327pp + sequence listing; English.  
 XX The invention relates to a spatially-addressable set of single exon  
 XX nucleic acid probes for measuring gene expression in a sample derived  
 XX from human breast and BF 474 cells. The method involves contacting  
 XX the probes with a collection of detectably labelled nucleic acids  
 XX derived from mRNA of human breast, and then measuring the label  
 XX bound to each probe of the microarray. The probes are useful for  
 XX verifying the expression of regions of genomic DNA predicted to  
 XX encode proteins. They are useful for gene discovery, and for  
 XX determining predisposition and/or prognosing breast disease. Gene  
 XX expression analysis is useful for assessing the toxicity of chemical  
 XX agents on cells. The microarray of this invention presents a far greater  
 XX diversity of probes for measuring gene expression, with far less bias  
 XX than expressed sequence tag microarrays. The method is suitable for  
 XX rapid production of functional information from genomic sequence. The  
 XX present sequence is a single exon nucleic acid probe of the invention.  
 XX Note: The sequence data for this patent did not form part of the  
 XX printed specification, but was obtained in electronic format directly  
 XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 479 BP; 126 A; 134 C; 82 G; 137 T; 0 other;

Query Match 33.6%; Score 24.2; DB 22; Length 479;  
 Best Local Similarity 59.4%; Pred. No. 7.4;  
 Matches 41; Conservative 0; Mismatches 28; Indels 0; Gaps 0;  
 Qy 2 ctttcaggcagatgcttcgggtccttaagacacaccccttctgtctgtcatgtgagc 61

Db 72 CTTTCAGAGCAGCAATGCTCCTGAGACACAGACCTTGCTATCACTAGCTGTGG 13  
QY 62 ccaacagt 70  
Db 12 CCTGAGAG 4

Search completed: July 31, 2002, 14:11:17  
Job time: 17374 sec

RESULT 15  
ABA54363/C  
ID ABA54363 standard; DNA; 479 BP.  
XX  
AC ABA54363;  
XX  
DT 01-FEB-2002 (first entry)  
XX  
DE Human foetal liver single exon nucleic acid probe #2668.  
XX  
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157277-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00669.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-483447/52.  
DR  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human fetal liver -  
XX  
PS Claim 1; SEQ ID NO 2668; 639pp + sequence listing; English.  
XX  
CC The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC foetal liver. The present sequence is a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 479 BP; 126 A; 134 C; 82 G; 137 T; 0 other;

Query Match 33.6%; Score 24.2; DB 22; Length 479;  
Best Local Similarity 59.4%; Pred. No. 7.4;  
Matches 41; Conservative 0; Mismatches 28; Indels 0; Gaps 0;  
QY 2 ctttcagacatgcttcggtgcttaagacacacaccttggctgtgtatgacatgtgagc 61  
Db 72 CTTTCAGAGCAGCAATGCTCCTGAGACACAGACCTTGCTATCACTAGCTGTGG 13  
QY 62 ccaacagt 70  
Db 12 CCTGAGAG 4









Query Match 100.0%; Score 72; DB 6; Length 72;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-16;  
 Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccttcagagacatgctgctgcttaagacacactcttgctgctatgacatgtgag 60  
 1 cctttcagagacatgctgctgcttaagacacactcttgctgctatgacatgtgag 60

DB 61 cccaacagtgc 72  
 61 cccaacagtgc 72

## RESULT 2

AX349063 3785 bp DNA linear PAT 06-FEB-2002  
 LOCUS AX349063  
 DEFINITION Sequence 1 from Patent WO0202785.  
 ACCESSION AX349063  
 VERSION AX349063.1 GI:18615098  
 KEYWORDS

## ORGANISM

SOURCE bread wheat.  
 Trilicium aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poidea; Triticeae; Triticum.

REFERENCE Sprunck, S., Kluth, A., Becker, D., Luetticke, S. and Loefer, H.  
 AUTHORS Promoters of gene expression in plant carypsops  
 TITLE Patent: WO 0202785-A 1 10-JAN-2002;  
 JOURNAL Aventis CropScience GmbH (DE)

## FEATURES

source 1..3785  
 /organism="Triticum aestivum"  
 /db\_xref="taxon:4565"  
 BASE COUNT 945 a 980 c 899 g 961 t  
 ORIGIN

Query Match 94.4%; Score 68; DB 6; Length 3785;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-15;  
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccttcagagacatgctgctgcttaagacacactcttgctgctatgacatgtgag 60  
 1 ccttcagagacatgctgctgcttaagacacactcttgctgctatgacatgtgag 60

DB 1936 ccttcagagacatgctgctgcttaagacacactcttgctgctatgacatgtgag 1995  
 61 cccaacagtgc 68  
 61 cccaacagtgc 68

DB 1996 cccaacagtgc 2003

RESULT 3  
 AP004636 147220 bp DNA linear HTG 17-JAN-2002  
 LOCUS AP004636  
 DEFINITION Oryza sativa chromosome 8 clone P0685B10, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, in ordered pieces.

ACCESSION AP004636  
 VERSION AP004636.1 GI:18182016  
 KEYWORDS HTG; HTGS; PHASE2.

SOURCE Oryza sativa (cultivar: Nipponbare) DNA, clone: P0685B10.  
 ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzae; Oryza.

REFERENCE 1 (bases 1 to 147220)  
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.

TITLE Direct Submission  
 JOURNAL Submitted (16-JAN-2002) Takuji Sasaki, National Institute of  
 Agricultural Sciences, Rice Genome Research Program, Kannondai  
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
 (E-mail: tsasaki@nias.affrc.go.jp, URL: http://rpg.dna.affrc.go.jp/,  
 Tel: 81-298-38-7441, Fax: 81-298-38-7468)

## COMMENT

NOTE: It currently consists of 1 contigs. Gaps between the contigs  
 are represented as runs of N. The order of the pieces is believed  
 to be correct as given, however the sizes of the gaps between them  
 are based on estimates that have provided by the submitter. This  
 sequence will be replaced by the finished sequence as soon as it is  
 available and the accession number will be preserved.  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

## FEATURES

## source

1..147220  
 /organism="Oryza sativa"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:4530"  
 /chromosome="8"  
 /clone="P0685B10"

BASE COUNT 4225 a 32162 c 31232 g 41469 t 102 others  
 ORIGIN

Query Match 40.8%; Score 29.4; DB 2; Length 147220;  
 Best Local Similarity 76.6%; Pred. No. 1.5;  
 Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 17 ttgcgtgcttaagacacactcttgctgctatgacatgtgagcc 63  
 17 ttgcgtgcttaagacacactcttgctgctatgacatgtgagcc 63

## RESULT 4

AX349063 3785 bp DNA linear PAT 06-FEB-2002  
 LOCUS AX349063  
 DEFINITION Sequence 1 from Patent WO0202785.  
 ACCESSION AX349063  
 VERSION AX349063.1 GI:18615098  
 KEYWORDS

## ORGANISM

SOURCE bread wheat.  
 Trilicium aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poidea; Triticeae; Triticum.

REFERENCE Sprunck, S., Kluth, A., Becker, D., Luetticke, S. and Loefer, H.  
 AUTHORS Promoters of gene expression in plant carypsops  
 TITLE Patent: WO 0202785-A 1 10-JAN-2002;  
 JOURNAL Aventis CropScience GmbH (DE)

## FEATURES

source 1..3785  
 /organism="Triticum aestivum"  
 /db\_xref="taxon:4565"  
 BASE COUNT 945 a 980 c 899 g 961 t  
 ORIGIN

Query Match 39.7%; Score 28.6; DB 6; Length 3785;  
 Best Local Similarity 76.2%; Pred. No. 3.4;  
 Matches 48; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 6 cagagacatgctgctgctt-aagacacactcttgctgctatgacatgtgagcca 64  
 6 cagagacatgctgctgctt-aagacacactcttgctgctatgacatgtgagcca 64

DB 2067 CGGAGCAGCAGCTTGGTGTAAAGGACCTGCTTGAATGACATGTGGCCAG 2008  
 65 aca 67  
 65 aca 67

## RESULT 5

AC008775 135503 bp DNA linear PRI 15-AUG-2001  
 LOCUS AC008775  
 DEFINITION Homo sapiens chromosome 5 clone CTD-2019N2, complete sequence.  
 ACCESSION AC008775



TITLE  
JOURNAL  
COMMENT

Campopiano, A., Castile, A., Choepe, Y., Colangelo, M., Collins, S.,  
Collimore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J. S.,  
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
Gajanan, J., Gardyna, S., Glende, S., Coyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,  
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, D.,  
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
Meldrum, J., Meneau, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, T. M., Olive, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Tessier, S., Theodore, J., Tirrell, A., Travers, M., Trigglio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (09-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 10, 2000 this sequence version replaced gi:8017214.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR

Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

Project Information

Center Project name: 528\_K\_12  
Center Clone name: 528\_K\_12

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731

Consensus quality: 182545 bases at least Q40

Consensus quality: 189073 bases at least Q30

Consensus quality: 192263 bases at least Q20

Insert size: 208000; agarose-fp

Insert size: 194499; sum-of-ctrls

Quality coverage: 4.3 in Q20 bases; agarose-fp

Quality coverage: 4.6 in Q20 bases; sum-of-ctrls

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 22 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 5530: contig of 5530 bp in length  
\* 5531 5630: gap of 100 bp  
\* 5631 7168: contig of 1538 bp in length  
\* 7169 7268: gap of 100 bp  
\* 7269 8574: contig of 1306 bp in length  
\* 8575 8674: gap of 100 bp  
\* 8675 9935: contig of 1261 bp in length  
\* 9936 10035: gap of 100 bp  
\* 10036 11861: contig of 1826 bp in length  
\* 11862 11961: gap of 100 bp  
\* 11962 14840: contig of 2879 bp in length  
\* 14841 14940: gap of 100 bp  
\* 14941 18392: contig of 3452 bp in length  
\* 18393 18492: gap of 100 bp  
\* 18493 22237: contig of 3745 bp in length  
\* 22238 22337: gap of 100 bp  
\* 22338 25371: contig of 3034 bp in length  
\* 25372 25472: gap of 100 bp  
\* 25473 31306: contig of 5835 bp in length  
\* 31307 31406: gap of 100 bp

## FEATURES

source

31407 36289: contig of 4883 bp in length  
\* 36290 36389: gap of 100 bp  
\* 36390 42370: contig of 5981 bp in length  
\* 42371 42470: gap of 100 bp  
\* 42471 50646: contig of 8176 bp in length  
\* 50647 50746: gap of 100 bp  
\* 50747 58275: contig of 7529 bp in length  
\* 58276 58375: gap of 100 bp  
\* 58376 69446: contig of 11071 bp in length  
\* 69447 69546: gap of 100 bp  
\* 69547 82175: contig of 12629 bp in length  
\* 82176 82275: gap of 100 bp  
\* 82276 108127: contig of 25852 bp in length  
\* 108128 108227: gap of 100 bp  
\* 108228 125265: contig of 17038 bp in length  
\* 125266 125365: gap of 100 bp  
\* 125366 145214: contig of 19849 bp in length  
\* 145215 145314: gap of 100 bp  
\* 145315 161113: contig of 15799 bp in length  
\* 161114 161213: gap of 100 bp  
\* 161214 183803: contig of 22590 bp in length  
\* 183804 183903: gap of 100 bp  
\* 183904 196599: contig of 12696 bp in length.  
Location/Qualifiers  
1. 196599  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-528K12"  
/clone\_11b="RPCT-11 Human Male BAC"  
1. 5530  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left"  
5631. 7168  
/note="assembly\_fragment"  
7269. 8574  
/note="assembly\_fragment"  
8675. 9935  
/note="assembly\_fragment"  
10036. 11861  
/note="assembly\_fragment"  
11962. 14840  
/note="assembly\_fragment"  
14941. 18392  
/note="assembly\_fragment"  
18493. 22237  
/note="assembly\_fragment"  
22338. 25371  
/note="assembly\_fragment"  
25472. 31306  
/note="assembly\_fragment"  
31407. 36289  
/note="assembly\_fragment"  
36390. 42370  
/note="assembly\_fragment"  
42471. 50646  
/note="assembly\_fragment"  
50747. 58275  
/note="assembly\_fragment"  
58376. 69446  
/note="assembly\_fragment"  
69547. 82175  
/note="assembly\_fragment"  
82276. 108127  
/note="assembly\_fragment"  
108228. 125265  
/note="assembly\_fragment"  
125366. 145214  
/note="assembly\_fragment"  
145315. 161113  
/note="assembly\_fragment"

misc\_feature 161214..183803  
/note="assembly\_fragment"  
misc\_feature 183904..196599  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right"

BASE COUNT 62063 a 36291 c 36295 g 59845 t 2105 others  
ORIGIN

Query Match 39.4%; Score 28.4; DB 2; Length 196599;  
Best Local Similarity 62.9%; Pred. No. 3.7;  
Matches 44; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

OY 2 cttcagagatgctgcgtgcttaagacactactgtgtgtcatgacatgagc 61  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 97617 ctttctgagatgctggtgagagacaaacatgaaatgagctgtctcaattgac 97558  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
OY 62 ccacacgtg 71  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 97557 TCACACTGC 97548

RESULT 8  
AC097952 168203 bp DNA linear HTG 20-DEC-2001  
LOCUS AC097952/c  
DEFINITION Rattus norvegicus clone CH230-127L5, \*\*\* SEQUENCING IN PROGRESS  
AC097952 63 unordered pieces.  
VERSION AC097952.2 GI:17948834  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 168203)

Alshrooke,S.L., Amarante,H.C., Are,J.R., Banks,T., Barbara,J.,  
Benton,J., Bimige,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
Burck,P., Burrell,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,  
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Escoto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,  
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Haves,A.,  
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,  
Hollins,B., Homel,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,  
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,  
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,  
Kovar,C., Kratoch,J., Kureshi,A., Landry,N., Ieal,B., Lewis,L.C.,  
Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,  
Luissegel,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,  
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,  
Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabadi,K.,  
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nokenko,S.,  
Ogih,M., Okunou,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,  
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
Rutledge,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,  
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,  
Slisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,  
Stone,H., Sutton,A., Swatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,  
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,  
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,  
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S.,  
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 168203)  
Worley,K.C.  
Submitted (23-OCT-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Dec 20, 2001 this sequence version replaced gi:16327658.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GFRW  
Center clone name: CH230-127L5  
----- Summary Statistics  
findphraplist  
Consensus quality: 134713 bases at least Q40  
Consensus quality: 142480 bases at least Q30  
Consensus quality: 148683 bases at least Q20  
Estimated insert size: 145761; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-ef estimation  
Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 63 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1  
5023: contig of 5023 bp in length  
5024 5123: gap of unknown length  
5124 11265: contig of 6142 bp in length  
11266 11365: gap of unknown length  
11366 17770: contig of 6405 bp in length  
17771 17870: gap of unknown length  
17871 22349: contig of 4479 bp in length  
22350 22449: gap of unknown length  
22450 28379: contig of 5830 bp in length  
28380 28379: gap of unknown length  
28380 32474: contig of 4095 bp in length  
32475 32574: gap of unknown length  
32575 36200: contig of 3626 bp in length  
36201 36300: gap of unknown length  
36301 39803: contig of 3503 bp in length  
39804 39903: gap of unknown length  
39904 41156: contig of 1253 bp in length  
41157 41256: gap of unknown length  
41257 44476: contig of 3220 bp in length  
44477 44576: gap of unknown length  
44577 47337: contig of 2761 bp in length  
47338 47437: gap of unknown length  
47438 51341: contig of 3904 bp in length  
51342 51441: gap of unknown length  
51442 54410: contig of 2969 bp in length  
54411 54510: gap of unknown length  
54511 58903: contig of 4393 bp in length  
58904 59003: gap of unknown length  
59004 61362: contig of 2359 bp in length  
61363 61462: gap of unknown length  
61463 65650: contig of 4188 bp in length  
65651 65750: gap of unknown length  
65751 70912: contig of 5162 bp in length







\* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 3720: contig of 3720 bp in length
3721 3820: gap of unknown length
3821 8996: contig of 5176 bp in length
8997 9096: gap of unknown length
9097 15347: contig of 6251 bp in length
15348 15447: gap of unknown length
15448 20807: contig of 5360 bp in length
20808 20907: gap of unknown length
20908 24965: contig of 4058 bp in length
24966 25065: gap of unknown length
25066 29243: contig of 4178 bp in length
29244 29343: gap of unknown length
29344 32951: contig of 3508 bp in length
32952 32951: gap of unknown length
32952 35458: contig of 2507 bp in length
35459 35558: gap of unknown length
35559 38502: contig of 2944 bp in length
38503 38602: gap of unknown length
38603 42551: contig of 3949 bp in length
42552 42551: gap of unknown length
42552 44528: contig of 1877 bp in length
44529 44628: gap of unknown length
44629 47051: contig of 2423 bp in length
47052 47151: gap of unknown length
47152 49622: contig of 2471 bp in length
49623 49722: gap of unknown length
49723 52560: contig of 2538 bp in length
52561 52660: gap of unknown length
52661 54288: contig of 1928 bp in length
54289 54388: gap of unknown length
54389 55786: contig of 1398 bp in length
55787 55886: gap of unknown length
55887 57012: contig of 1126 bp in length
57013 57112: gap of unknown length
57113 58518: contig of 1406 bp in length
58519 58618: gap of unknown length
58619 59973: contig of 1355 bp in length
59974 60073: gap of unknown length
60074 61465: contig of 1392 bp in length
61466 61565: gap of unknown length
61566 62795: contig of 1230 bp in length
62796 62895: gap of unknown length
62896 64360: contig of 1465 bp in length
64361 64460: gap of unknown length
64461 66972: contig of 2512 bp in length
66973 67072: gap of unknown length
67073 68530: contig of 1458 bp in length
68531 68630: gap of unknown length
68631 69881: contig of 1251 bp in length
69882 71210: gap of unknown length
71211 71310: contig of 1229 bp in length
71311 73418: gap of unknown length
73419 73518: contig of 2108 bp in length
73519 75998: gap of unknown length
75999 76098: gap of unknown length
76099 77417: contig of 1319 bp in length
77418 77518: gap of unknown length
77519 79011: contig of 1394 bp in length
79012 80270: gap of unknown length
80271 80370: contig of 1259 bp in length
80371 81395: gap of unknown length
81396 81495: contig of 1025 bp in length
81496 81595: gap of unknown length
81596 83252: contig of 1657 bp in length
83253 84749: gap of unknown length
84750 84849: gap of unknown length

```

```

* 84850 86157: contig of 1308 bp in length
* 86158 86257: gap of unknown length
* 86258 87827: contig of 1570 bp in length
* 87828 87927: gap of unknown length
* 87928 89649: contig of 1722 bp in length
* 89650 89749: gap of unknown length
* 89750 91608: contig of 1859 bp in length
* 91609 91708: gap of unknown length
* 91709 92881: contig of 1173 bp in length
* 92882 92981: gap of unknown length
* 92982 94509: contig of 1528 bp in length
* 94510 94609: gap of unknown length
* 94610 95936: contig of 1327 bp in length
* 95937 96036: gap of unknown length
* 96037 97129: contig of 1093 bp in length
* 97130 97229: gap of unknown length
* 97230 98432: contig of 1203 bp in length
* 98433 98532: gap of unknown length
* 98533 99710: contig of 1178 bp in length
* 99711 99810: gap of unknown length
* 99811 100929: contig of 1119 bp in length
* 100930 101029: gap of unknown length
* 101030 102053: contig of 1024 bp in length.

```

FEATURES  
 source 1..102053  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 /clone="CH230-8N7"

BASE COUNT 27421 a 20865 c 20727 g 28460 t 4580 others

ORIGIN

Query Match 37.5%; Score 27; DB 2; Length 102053;  
 Best Local Similarity 62.7%; Pred. No. 13; Indels 0; Gaps 0;  
 Matches 42; Conservative 0; Mismatches 25;

```

QY 4 ttccagcagatcttcgctgccttaagacacacaccttgctctatcacatgagccc 63
Db 20796 TTCCAGTGATGCTAGGTCCTTATGAAATGATTCACCTACCTATGATGCC 20737
QY 64 aacagatg 70
Db 20736 AACACTG 20730

```

RESULT 12  
 AP003236 167399 bp DNA linear PLN 08-SEP-2001

LOCUS  
 DEFINITION  
 Oryza sativa genomic DNA, chromosome 1, PAC clone: P0043B10.

AP003236  
 AP003236.3 GI:15528647

KEYWORDS  
 Oryza sativa (cultivar: Nipponbare) DNA, clone: P0043B10.

ORGANISM  
 Oryza sativa  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Enthartoideae; Oryzaceae; Oryza.

REFERENCE  
 1 (bases 1 to 167399)  
 Sasaki, T., Matsumoto, T. and Yamamoto, K.

Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
 clone: P0043B10

Published Only in Database (2001) In press

2 (bases 1 to 167399)

Sasaki, T., Matsumoto, T. and Yamamoto, K.

Direct Submission  
 Submitted (19-FEB-2001) Takuji Sasaki, National Institute of  
 Agrobiological Sciences, Rice Genome Research Program, Kannondai

2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
 (E-mail: sasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/.

Tel: 81-298-38-7441, Fax: 81-298-38-7468.

On Sep 7, 2001 this sequence version replaced gi:14149142.

Genes were predicted from the integrated results of the following:

GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor







## COMMENT

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Nov 15, 2001 this sequence version replaced gi:15625026.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)

----- Summary Statistics

Center project name: H\_NH0811J09

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University) and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateo, M., Catanese, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://Daepac.med.buffalo.edu>)

VECTOR: pBAC3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-563K23; the clone sequenced to the right is RP11-298A10, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-811J9; actual end is at base position 140619 of RP11-298A10.

Sequence derived from PCR, base position 6122 to 6452 and 7046 to 7087.

## FEATURES

Polymorphisms have been identified between AC073264 and AC092214.

## Location/Qualifiers

1. .72045

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="7"

/map="7"

/clone="RP11-811J9"

/clone\_lib="RPCI-11"

1. .437

/note="match to EST A1830201 (NID:g5450861) wj78e05.x1"

1. .333

/note="match to EST BG777585 (NID:g14047902)"

misc\_feature

1. .169

/note="match to EST A1066553 (NID:g3367255) ool4902.x1"

125. .1440

/note="CPG\_island (%GC=73.4, o/e=0.90, #CGs=130)"

357. .360

/note="match to EST BF733203 (NID:g10147195)"

444. .573

misc\_feature

misc\_feature

/note="similar to EST BM049883 (NID:g16779150)"  
481. .482

misc\_feature

/note="similar to Homo sapiens EST BG327560 (NID:g131333988)"

misc\_feature

/note="match to EST BF205674 (NID:g11099260)"

misc\_feature

651. .678

/note="similar to Homo sapiens EST BE273823 (NID:g9148759)"

misc\_feature

673. .773

/note="match to EST B1522595 (NID:g15347387)"

misc\_feature

683. .773

/note="match to EST AL556892 (NID:g12899973)"

misc\_feature

688. .773

/note="similar to Mus musculus EST BB619995 (NID:g16459189)"

misc\_feature

691

/note="match to EST B1833448 (NID:g15944998)"

misc\_feature

692. .773

/note="match to EST B1826228 (NID:g15937778)"

misc\_feature

692. .742

/note="similar to Homo sapiens EST B1768766 (NID:g15760344)"

misc\_feature

699. .773

/note="match to EST AL545268 (NID:g12877749)"

misc\_feature

724. .773

/note="similar to Homo sapiens EST B1827429 (NID:g15938979)"

misc\_feature

735. .773

/note="match to EST B1833448 (NID:g15944998)"

misc\_feature

738. .773

/note="similar to Homo sapiens EST BG327560 (NID:g131333988)"

misc\_feature

778. .837

/note="similar to Homo sapiens EST BF528670 (NID:g11616033)"

repeat\_region

778. .837

/note="similar to Homo sapiens EST BG327560 (NID:g131333988)"

misc\_feature

862. .1133

/note="match to EST B1457310 (NID:g15247966)"

misc\_feature

863. .1001

/note="match to EST A292501 (NID:g1940488) zt30b03.r1"

misc\_feature

879. .1143

/note="match to EST BG831833 (NID:g14179420)"

misc\_feature

921. .1143

/note="match to EST B1833448 (NID:g15944998)"

misc\_feature

921. .1127

/note="similar to Homo sapiens EST B1827429 (NID:g15938979)"

misc\_feature

921. .1143

/note="similar to Homo sapiens EST BG327560 (NID:g131333988)"

misc\_feature

921. .1143

/note="match to EST B1826228 (NID:g15937778)"

misc\_feature

923. .1143

/note="similar to Homo sapiens EST BF311877 (NID:g9148556)"

misc\_feature

923. .1143

/note="similar to Mus musculus EST AM229324 (NID:g5558620) uo09c03.y1"

misc\_feature

924. .1143

/note="match to EST AL545268 (NID:g12877749)"

misc\_feature

924. .1139

/note="similar to Homo sapiens EST BF528670 (NID:g11616033)"

misc\_feature

924. .1142

/note="match to EST B1522595 (NID:g15347387)"

misc\_feature

924. .1143

/note="match to EST AL556892 (NID:g12899973)"

misc\_feature

924. .1143

/note="similar to Mus musculus EST BB619995 (NID:g16459189)"

misc\_feature

939. .1143



```

misc_feature      /note="assembly_fragment"
84154..92774      /note="assembly_fragment"
misc_feature      92875..100657
100758..107529    /note="assembly_fragment"
misc_feature      /note="assembly_fragment"
107630..113210    /note="assembly_fragment"
misc_feature      113311..117568
117669..120668    /note="assembly_fragment clone_end:SP6 vector_side:left"
misc_feature      /note="assembly_fragment"
120769..122717    /note="assembly_fragment"
misc_feature      /note="assembly_fragment"
BASE COUNT      37157 a 24437 c 24619 g 35304 t 1200 others
ORIGIN

```

```

Query Match      37.2%: Score 26.8; DB 2; Length 122717;
Best Local Similarity 68.5%; Pred. No. 15;
Matches 37; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
Oy 17 ttctgtgccttaagacacctcttctgtctatgacatgtgagccacacgtg 70
      |||||  |||  |||||  |||||  |||||  |||||  |||||  |||||
Db 82525 TTtGTACCCCAAACTCTACCTTATTGTTATTTCAAAAGAGCCAAAAGTG 82578

```

Search completed: July 31, 2002, 14:01:44  
 Job time: 17546 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:20:34 ; Search time 165.21 Seconds  
(without alignments)  
31.223 Million cell updates/sec

Title: US-09-899-718A-3

Perfect score: 21

Sequence: 1 agaacaacaacaacaaca 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 21    | 100.0       | 2224   | 4     | US-08-477-347-2    |
| 2          | 21    | 100.0       | 2224   | 4     | US-08-476-862-1    |
| 3          | 21    | 100.0       | 11811  | 4     | US-09-078-294-7    |
| 4          | 20    | 95.2        | 369    | 4     | US-08-991-789A-191 |
| 5          | 20    | 95.2        | 1265   | 4     | US-09-062-451-191  |
| 6          | 20    | 95.2        | 1265   | 4     | US-08-991-789A-169 |
| 7          | 20    | 95.2        | 1265   | 4     | US-09-062-451-169  |
| 8          | 20    | 95.2        | 1646   | 4     | US-08-991-789A-290 |
| 9          | 20    | 95.2        | 1646   | 4     | US-09-062-451-290  |
| 10         | 19.4  | 92.4        | 500    | 2     | US-08-967-101-86   |
| 11         | 19.4  | 92.4        | 500    | 2     | US-08-592-541-86   |
| 12         | 19.4  | 92.4        | 500    | 3     | US-09-124-698-86   |
| 13         | 19.4  | 92.4        | 500    | 4     | US-09-127-480-86   |
| 14         | 19.4  | 92.4        | 500    | 4     | US-08-496-841C-86  |
| 15         | 19.4  | 92.4        | 504    | 4     | US-09-328-111-120  |
| 16         | 19.4  | 92.4        | 550    | 4     | US-09-306-042-2    |
| 17         | 19.4  | 92.4        | 678    | 4     | US-09-227-357-81   |
| 18         | 19.4  | 92.4        | 1140   | 4     | US-08-943-731-209  |
| 19         | 19.4  | 92.4        | 1220   | 4     | US-09-227-357-54   |
| 20         | 19.4  | 92.4        | 1300   | 2     | US-08-474-020A-13  |
| 21         | 19.4  | 92.4        | 1355   | 1     | US-07-757-390-4    |
| 22         | 19.4  | 92.4        | 1355   | 1     | US-07-757-390-4    |
| 23         | 19.4  | 92.4        | 1355   | 1     | US-08-442-282-4    |
| 24         | 19.4  | 92.4        | 1355   | 1     | US-08-442-282-4    |
| 25         | 19.4  | 92.4        | 1355   | 1     | US-08-442-281-4    |
| 26         | 19.4  | 92.4        | 1355   | 1     | US-08-442-281-16   |
| 27         | 19.4  | 92.4        | 1355   | 2     | US-08-939-727-4    |

|    |      |      |      |   |                   |                   |
|----|------|------|------|---|-------------------|-------------------|
| 28 | 19.4 | 92.4 | 1355 | 2 | US-08-939-727-16  | Sequence 16, Appl |
| 29 | 19.4 | 92.4 | 1390 | 4 | US-09-210-168-1   | Sequence 1, Appl  |
| 30 | 19.4 | 92.4 | 1506 | 3 | US-09-176-657-5   | Sequence 5, Appl  |
| 31 | 19.4 | 92.4 | 1641 | 4 | US-09-018-635-22  | Sequence 22, Appl |
| 32 | 19.4 | 92.4 | 1808 | 1 | US-07-757-390-2   | Sequence 2, Appl  |
| 33 | 19.4 | 92.4 | 1808 | 1 | US-07-757-390-15  | Sequence 15, Appl |
| 34 | 19.4 | 92.4 | 1808 | 1 | US-08-442-282-2   | Sequence 2, Appl  |
| 35 | 19.4 | 92.4 | 1808 | 1 | US-08-442-282-15  | Sequence 15, Appl |
| 36 | 19.4 | 92.4 | 1808 | 1 | US-08-442-281-2   | Sequence 2, Appl  |
| 37 | 19.4 | 92.4 | 1808 | 1 | US-08-442-281-15  | Sequence 15, Appl |
| 38 | 19.4 | 92.4 | 1808 | 2 | US-08-939-727-2   | Sequence 2, Appl  |
| 39 | 19.4 | 92.4 | 1808 | 2 | US-08-939-727-15  | Sequence 15, Appl |
| 40 | 19.4 | 92.4 | 1901 | 5 | PCT-US93-05000-32 | Sequence 32, Appl |
| 41 | 19.4 | 92.4 | 2022 | 2 | US-08-464-517-32  | Sequence 32, Appl |
| 42 | 19.4 | 92.4 | 2022 | 2 | US-08-246-361A-32 | Sequence 32, Appl |
| 43 | 19.4 | 92.4 | 2022 | 3 | US-08-463-772-32  | Sequence 32, Appl |
| 44 | 19.4 | 92.4 | 2035 | 3 | US-09-022-699-3   | Sequence 3, Appl  |
| 45 | 19.4 | 92.4 | 2192 | 1 | US-08-035-392-1   | Sequence 1, Appl  |

## ALIGNMENTS

RESULT 1  
US-08-477-347-2/c  
Sequence 2, Application US/08477347  
Patent No. 623246  
GENERAL INFORMATION:  
APPLICANT: WALLACH, David  
APPLICANT: BIGDA, Jacek  
APPLICANT: BELETSKY, Igor  
APPLICANT: METT, Igor  
TITLE OF INVENTION: TNF LIGANDS  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROMDY AND NEIMARK  
STREET: 419 Seventh Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,347  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/115,685  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 106271  
FILING DATE: 08-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Townsend, G. Kevin  
REGISTRATION NUMBER: 34,033  
REFERENCE/DOCKET NUMBER: WALLACH-10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2224 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS

LOCATION: 90..1472  
US-08-477-347-2

Query Match 100.0%; Score 21; DB 4; Length 2224;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaacaa 21  
|||||  
DB 1847 AGAAACAACAACAACAAA 1827

## RESULT 2

US-08-476-862-1/c  
Sequence 1, Application US/08476862  
Patent No. 6262239

## GENERAL INFORMATION:

APPLICANT: WALLACH, David  
APPLICANT: BIGDA, Jacek  
APPLICANT: BELETSKY, Igor  
APPLICANT: METT, Igor  
APPLICANT: ENGELMANN, Hartmut  
TITLE OF INVENTION: TNF INHIBITORS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/476,862  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: IL 107267  
FILING DATE: 12-OCT-1993

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: IL 94039  
FILING DATE: 06-APR-1990

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: IL 91229  
FILING DATE: 06-AUG-1989

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: IL 90339  
FILING DATE: 18-MAY-1989

## ATTORNEY/AGENT INFORMATION:

NAME: BROWDY, Roger L.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: WALLACH-12A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2224 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 90..1472

US-08-476-862-1

Query Match 100.0%; Score 21; DB 4; Length 2224;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaacaa 21  
|||||  
DB 1847 AGAAACAACAACAACAAA 1827

## RESULT 3

US-09-078-294-7/c  
Sequence 7, Application US/09078294  
Patent No. 6265211

## GENERAL INFORMATION:

APPLICANT: Choo, Kong-Hong Andy  
APPLICANT: Du Saitl, Desiree  
APPLICANT: Cancilla, Michael R.  
TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE  
FILE REFERENCE: Davies Col  
CURRENT APPLICATION NUMBER: US/09/078,294  
CURRENT FILING DATE: 1998-05-13  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 7  
LENGTH: 11811  
TYPE: DNA  
ORGANISM: BAC-F2 contig 3  
US-09-078-294-7

Query Match 100.0%; Score 21; DB 4; Length 11811;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaacaa 21  
|||||  
DB 3250 AGAAACAACAACAACAAA 3230

## RESULT 4

US-08-991-789A-191  
Sequence 191, Application US/08991789A  
Patent No. 6225054

## GENERAL INFORMATION:

APPLICANT: Frudakis, Tony N.

Reed, Steven G.

Smith, John M.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE

TREATMENT AND DIAGNOSIS OF BREAST CANCER

NUMBER OF SEQUENCES: 292

CORRESPONDENCE ADDRESS:

ADDRESSEE: Seed IP Law Group

STREET: 701 Fifth Avenue, Suite 6300

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/991,789A

FILING DATE: 11-Dec-1997

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Potter, Jane E. R.

REGISTRATION NUMBER: 33,332

REFERENCE/DOCKET NUMBER: 210121.419C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 662-6031

INFORMATION FOR SEQ ID NO: 191;  
SEQUENCE CHARACTERISTICS:  
LENGTH: 369 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 191:  
US-08-991-789A-191

Query Match 95.2%; Score 20; DB 4; Length 369;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaacaacaa 20  
|||||  
DB 148 AGAACAACAACAACAACA 167

RESULT 5  
US-09-062-451-191  
Sequence 191, Application US/09062451  
Patent No. 6344550

GENERAL INFORMATION:  
APPLICANT: Frudakis, Tony N.  
APPLICANT: Smith, John M.  
APPLICANT: Reed, Steven G.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
NUMBER OF SEQUENCES: 297  
TREATMENT AND DIAGNOSIS OF BREAST CANCER

CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/062,451  
FILING DATE: 04-APR-1997  
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:  
NAME: Mark, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.419C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 191:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 369 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-062-451-191

Query Match 95.2%; Score 20; DB 4; Length 369;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaacaacaa 20  
|||||  
DB 148 AGAACAACAACAACAACA 167

RESULT 6  
US-08-991-789A-169/c

Sequence 169, Application US/08991789A  
Patent No. 6225054  
GENERAL INFORMATION:  
APPLICANT: Frudakis, Tony N.  
APPLICANT: Smith, John M.  
APPLICANT: Reed, Steven G.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
TREATMENT AND DIAGNOSIS OF BREAST CANCER  
NUMBER OF SEQUENCES: 292  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seed IP Law Group  
STREET: 701 Fifth Avenue, Suite 6300  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/991,789A  
FILING DATE: 11-Dec-1997  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Potter, Jane E. R.  
REGISTRATION NUMBER: 33,332  
REFERENCE/DOCKET NUMBER: 210121.419C3

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 169:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1265 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 169:  
US-08-991-789A-169

Query Match 95.2%; Score 20; DB 4; Length 1265;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaacaacaa 20  
|||||  
DB 840 AGAACAACAACAACAACA 821

RESULT 7  
US-09-062-451-169/c  
Sequence 169, Application US/09062451  
Patent No. 6344550

GENERAL INFORMATION:  
APPLICANT: Frudakis, Tony N.  
APPLICANT: Smith, John M.  
APPLICANT: Reed, Steven G.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
TREATMENT AND DIAGNOSIS OF BREAST CANCER  
NUMBER OF SEQUENCES: 297  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/062,451  
FILING DATE: 04-APR-1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: MAKI, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.419C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 169:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1265 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-062-451-169

Query Match 95.2%; Score 20; DB 4; Length 1265;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaaca 20  
|||||  
Db 840 AGAAACAACAACAACA 821

RESULT 8  
US-08-991-789A-290/c  
Sequence 290, Application US/08991789A  
Patent No. 6225054  
GENERAL INFORMATION:  
APPLICANT: Frudakis, Tony N.  
Smith, John M.  
Reed, Steven G.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
TREATMENT AND DIAGNOSIS OF BREAST CANCER  
NUMBER OF SEQUENCES: 292  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seed IP Law Group  
STREET: 701 Fifth Avenue, Suite 6300  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/991,789A  
FILING DATE: 11-Dec-1997  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Potter, Jane E., R.  
REGISTRATION NUMBER: 33,332  
REFERENCE/DOCKET NUMBER: 210121.419C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 290:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1646 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 290:  
US-08-991-789A-290

Query Match 95.2%; Score 20; DB 4; Length 1646;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaaca 20  
|||||  
Db 1213 AGAAACAACAACAACA 1194

RESULT 9  
US-09-062-451-290/c  
Sequence 290, Application US/09062451  
Patent No. 6344550  
GENERAL INFORMATION:  
APPLICANT: Frudakis, Tony N.  
Smith, John M.  
Reed, Steven G.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
TREATMENT AND DIAGNOSIS OF BREAST CANCER  
NUMBER OF SEQUENCES: 297  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/062,451  
FILING DATE: 04-APR-1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: MAKI, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.419C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 290:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1646 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-062-451-290

Query Match 95.2%; Score 20; DB 4; Length 1646;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaaca 20  
|||||  
Db 1213 AGAAACAACAACAACA 1194

RESULT 10  
US-08-967-101-86  
Sequence 86, Application US/08967101  
Patent No. 5840540  
GENERAL INFORMATION:  
APPLICANT: ST. GEORGE-HYSLOP, PETER H  
ROMMENS, JOHANNA M  
FRASER, PAUL E  
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
TO ALZHEIMER'S DISEASE  
NUMBER OF SEQUENCES: 183

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
;; STREET: High Street Tower - 125 High Street  
;; CITY: Boston  
;; STATE: Massachusetts  
;; COUNTRY: U.S.A.  
;; ZIP: 02110  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/967,101  
;; FILING DATE: 10-NOV-1997  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/592,541  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Pitcher, Edmund R.  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 248-7000  
;; TELEFAX: (617) 248-7100  
;; INFORMATION FOR SEQ ID NO: 86:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 500 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; US-08-967-101-86

Query Match 92.4%; Score 19.4; DB 2; Length 500;  
Best Local Similarity 95.2%; Pred. No. 38;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 agaacaacaacaacaaca 21  
Db 70 ACAACAACAACAACAACA 90

RESULT 11  
US-08-592-541-86  
; Sequence 86, Application US/08592541  
; Patent No. 5986054  
; GENERAL INFORMATION:  
; APPLICANT: ST. GEORGE-HYSLOP, PETER H  
; APPLICANT: FRASER, PAUL E  
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE  
; NUMBER OF SEQUENCES: 183  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
; STREET: High Street Tower - 125 High Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/592,541  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pitcher, Edmund R.  
; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (617) 248-7000  
;; TELEFAX: (617) 248-7100  
;; INFORMATION FOR SEQ ID NO: 86:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 500 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; US-08-592-541-86

Query Match 92.4%; Score 19.4; DB 2; Length 500;  
Best Local Similarity 95.2%; Pred. No. 38;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 agaacaacaacaacaaca 21  
Db 70 ACAACAACAACAACAACA 90

RESULT 12  
US-09-124-698-86  
; Sequence 86, Application US/09124698  
; Patent No. 6117978  
; GENERAL INFORMATION:  
; APPLICANT: ST. GEORGE-HYSLOP, PETER H  
; APPLICANT: ROMMENS, JOHANNA M  
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE  
; NUMBER OF SEQUENCES: 183  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
; STREET: High Street Tower - 125 High Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/124,698  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/592,541  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pitcher, Edmund R.  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 86:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 500 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-09-124-698-86

Query Match 92.4%; Score 19.4; DB 3; Length 500;  
Best Local Similarity 95.2%; Pred. No. 38;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 agaacaacaacaacaaca 21  
Db 70 ACAACAACAACAACAACA 90

```

RESULT 13
US-09-127-480-86
; Sequence 86, Application US/09127480
; Patent No. 6194153
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HORWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/127,480
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,541
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-127-480-86

Query Match          92.4%; Score 19.4; DB 4; Length 500;
Best Local Similarity 95.2%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy      1   agaacacaacacaacaa 21
        | ||||| ||||| |||||
Db      70   ACAACCAACAACAACAACA 90

RESULT 14
US-08-496-841C-86
; Sequence 86, Application US/08496841C
; Patent No. 6210919
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby, PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
```

```

      ZIP: 10022
      COMPUTER READABLE FORM:
      MEDIUM TYPE: floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patentin Release #1.0, Version #1.30
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/496,841C
      FILING DATE: 28-Jun-1995
      CLASSIFICATION: <unknown>
      ATTORNEY/AGENT INFORMATION:
      NAME: Paul F. Fehner, Ph.D.
      REGISTRATION NUMBER: 35,135
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (212) 527-7700
      TELEFAX: (212) 753-6237
      INFORMATION FOR SEQ ID NO: 86:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 500 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: DNA (genomic)
      SEQUENCE DESCRIPTION: SEQ ID NO: 86:
      US-08-496-841C-86

      Query Match          92.4%; Score 19.4; DB 4; Length 500;
      Best Local Similarity 95.2%; Pred. No. 38;
      Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0.

      QY      1 agaacaacaacaacaaca 21
              | | | | | | | | | | | | | |
      Db      70 ACAACACAAACAACAACA 90

      RESULT 15
      US-09-328-111-120/c
      ; Sequence 120, Application US/09328111
      ; Patent No. 6262333
      ; GENERAL INFORMATION:
      ; APPLICANT: Endege, Wilson O.
      ; APPLICANT: Steinmann, Kathleen E.
      ; APPLICANT: Astle, Jon H.
      ; APPLICANT: Burgess, Christopher C.
      ; APPLICANT: Bushnell, Steven E.
      ; APPLICANT: Carroll III, Eddie
      ; APPLICANT: Catino, Theodore J.
      ; APPLICANT: Derti, Adnan
      ; APPLICANT: Ford, Donna M.
      ; APPLICANT: Lewis, Marcia E.
      ; APPLICANT: Monahan, John E.
      ; APPLICANT: Schlegel, Robert
      ; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
      ; FILE REFERENCE: CCD-257 (US)
      ; CURRENT APPLICATION NUMBER: US/09/328,111
      ; CURRENT FILING DATE: 1999-06-08
      ; EARLIER APPLICATION NUMBER: US 60/088,801
      ; EARLIER FILING DATE: 1998-06-10
      ; NUMBER OF SEQ ID NOS: 850
      ; SOFTWARE: FastSeq for Windows Version 3.0
      ; SEQ ID NO 120
      ; LENGTH: 504
      ; TYPE: DNA
      ; ORGANISM: Homo sapiens
      ; FEATURE:
      ; NAME/KEY: misc_feature
      ; LOCATION: (1)..(504)
      ; OTHER INFORMATION: n = A,T,C or G
      US-09-328-111-120
  
```

Thu Aug 1 08:29:59 2002

us-09-899-718a-3.rni

Page 7

|    | Query Match           | 92.4%                | Score 19.4   | DB 4 | Length 504 |
|----|-----------------------|----------------------|--------------|------|------------|
|    | Best Local Similarity | 95.2%                | Pred. No. 38 |      |            |
|    | Matches               | 20                   | Conservative | 0    | Mismatches |
|    |                       |                      |              | 1    | Indels     |
|    |                       |                      |              |      | Gaps       |
| Oy | 1                     | agaacaaacaaacaaacaa  | 21           |      |            |
|    |                       |                      |              |      |            |
| Db | 402                   | AAAAACAAACAAACAAACAA | 382          |      |            |

Search completed: July 31, 2002, 12:20:35  
Job time: 11658 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:11:53 ; Search time 5855.71 Seconds  
(Without alignments)  
48.403 Million cell updates/sec

Title: US-09-899-718a-3  
Perfect score: 21  
Sequence: 1 agaacaacaacaacaaca 21

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estlba:\*  
2: em\_estlhum:\*  
3: em\_estlin:\*  
4: em\_estlm:\*  
5: em\_estm:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| c 1        | 21    | 100.0       | 58     | 12 | AZ504073 1M0344C07 |
| c 2        | 21    | 100.0       | 100    | 10 | B1046075           |
| c 3        | 21    | 100.0       | 139    | 10 | BG988551           |
| c 4        | 21    | 100.0       | 144    | 10 | B1043282           |
| c 5        | 21    | 100.0       | 148    | 9  | A1345073           |
| c 6        | 21    | 100.0       | 149    | 10 | BF922085           |
| c 7        | 21    | 100.0       | 150    | 9  | AA639520           |
| c 8        | 21    | 100.0       | 156    | 12 | A2493323           |
| c 9        | 21    | 100.0       | 156    | 12 | BH105347           |
| c 10       | 21    | 100.0       | 167    | 9  | A1253971           |
| c 11       | 21    | 100.0       | 167    | 9  | A1733948           |
| c 12       | 21    | 100.0       | 177    | 9  | AA719825           |
| c 13       | 21    | 100.0       | 180    | 9  | AA457363           |
| c 14       | 21    | 100.0       | 181    | 9  | A1733998           |
| c 15       | 21    | 100.0       | 184    | 9  | BB072882           |
| c 16       | 21    | 100.0       | 187    | 9  | A1251238           |
| c 17       | 21    | 100.0       | 187    | 9  | A1792899           |

| c    | 18 | 21    | 100.0 | 190 | 12       | AZ504073 |
|------|----|-------|-------|-----|----------|----------|
| c 19 | 21 | 100.0 | 194   | 9   | A1549855 |          |
| c 20 | 21 | 100.0 | 196   | 9   | A1792857 |          |
| c 21 | 21 | 100.0 | 201   | 9   | AA839213 |          |
| c 22 | 21 | 100.0 | 202   | 9   | A1465751 |          |
| c 23 | 21 | 100.0 | 222   | 9   | BB150403 |          |
| c 24 | 21 | 100.0 | 223   | 9   | BB536522 |          |
| c 25 | 21 | 100.0 | 224   | 9   | BE071169 |          |
| c 26 | 21 | 100.0 | 227   | 9   | AV317088 |          |
| c 27 | 21 | 100.0 | 228   | 9   | BB183787 |          |
| c 28 | 21 | 100.0 | 231   | 9   | AV229441 |          |
| c 29 | 21 | 100.0 | 239   | 10  | BF747295 |          |
| c 30 | 21 | 100.0 | 241   | 12  | A2221788 |          |
| c 31 | 21 | 100.0 | 243   | 9   | AV281580 |          |
| c 32 | 21 | 100.0 | 245   | 9   | AV262627 |          |
| c 33 | 21 | 100.0 | 247   | 9   | AV230114 |          |
| c 34 | 21 | 100.0 | 249   | 9   | BB852501 |          |
| c 35 | 21 | 100.0 | 251   | 9   | AV336252 |          |
| c 36 | 21 | 100.0 | 255   | 12  | A2414740 |          |
| c 37 | 21 | 100.0 | 256   | 9   | AV264285 |          |
| c 38 | 21 | 100.0 | 257   | 10  | T73967   |          |
| c 39 | 21 | 100.0 | 260   | 12  | A2597098 |          |
| c 40 | 21 | 100.0 | 262   | 9   | BB425056 |          |
| c 41 | 21 | 100.0 | 264   | 10  | BB850705 |          |
| c 42 | 21 | 100.0 | 267   | 12  | A0675744 |          |
| c 43 | 21 | 100.0 | 268   | 12  | A2413748 |          |
| c 44 | 21 | 100.0 | 270   | 12  | A2772597 |          |
| c 45 | 21 | 100.0 | 275   | 9   | BB482086 |          |

#### ALIGNMENTS

RESULT 1  
AZ504073/c 58 bp DNA linear GSS 05-Oct-2000  
DEFINITION  
LOCUS  
1M0344C07F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0344C07 F, DNA sequence.

ACCESSION  
AZ504073  
VERSION  
AZ504073.1 GI:10685389

KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

TITLE

JOURNAL  
COMMENT

FEATURES  
source

Mouse whole genome scaffolding with paired end reads from 10kb  
Plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: dduan@genetics.utah.edu  
Insert length: 10000 Std Error: 0.00  
Plate: 0344 row: C column: 07  
Seg primer: CGTTTAAACGACGCCAGT  
Class: plasmid ends  
High quality sequence stop: 58.  
Location/Qualifiers  
1..58  
/organism="Mus musculus"  
/strain="CS7BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0344C07"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"

```

/ssex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/Note="Vector: pMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g1147321141gb/AT129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT      11 a          9 c          10 g          28 t
ORIGIN

Query Match      100.0%; Score 21; DB 12; Length 58;
Best Local Similarity 100.0%; Pred. No. 9.6e+04;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 agaaacaacaacaacaaca 21
Db 23 AGAAACAACAACAACAACA 3

RESULT 2
LOCUS      BJ046075      100 bp      mRNA      linear      EST 06-DEC-2001
DEFINITION laevis cDNA clone XL00602 3', mRNA sequence.
ACCESSION  BJ046075
VERSION     BJ046075.1 GI:17398266
KEYWORDS    African clawed frog.
SOURCE      Xenopus laevis
ORGANISM    Xenopus laevis
REFERENCE   1 (bases 1 to 100)
AUTHORS    Kitayama,A., Terasaka,C., Mochi,M., Ueno,N., Shin-I,T. and Kohara
Y.
TITLE      Expressed genes in X. laevis embryo
JOURNAL    Unpublished (2001)
COMMENT    Contact: Tadashi Shin-I
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshin@genes.nig.ac.jp.
FEATURES
source      1..100
            /organism="Xenopus laevis"
            /db_xref="taxon:8355"
            /clone="XL00602"
            /clone_lib="NIBB Mochi normalized Xenopus neurula
            library"
            /tissue_type="whole embryo"
            /dev_stage="stage 15"
BASE COUNT  42 a          12 c          12 g          32 t          2 others
ORIGIN

Query Match      100.0%; Score 21; DB 10; Length 100;

```

```

Best Local Similarity 100.0%; Pred. No. 7.3e+04;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 agaaacaacaacaacaaca 21
Db 22 AGAAACAACAACAACAACA 42

RESULT 3
LOCUS      BG998551      139 bp      mRNA      linear      EST 13-JUN-2001
DEFINITION PM4-HT1301-180401-003-g05_1 HT1301 Homo sapiens cDNA, mRNA
sequence.
ACCESSION  BG998551
VERSION     BG998551.1 GI:14402621
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 139)
AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Bioness,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            M., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE    20202663
COMMENT    Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM4&t=PM4-HT1301-
            180401-003-g05_1&t3=2001-04-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 139.
FEATURES
source      1..139
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="HT1301"
            /dev_stage="Adult"
            /Note="Organ: head_neck; Vector: puc18; Site_1: Sma1;
            Site_2: Sma1; A mini-library was made by cloning products
            derived from ORESTES PCR (U.S. Letters Patent application
            No. 196,716 - Ludwig Institute for Cancer Research)
            profiles into the puc 18 vector. Reverse transcription of
            tissue mRNA and cDNA amplification were performed under
            low stringency conditions."
BASE COUNT  54 a          29 c          20 g          20 t
ORIGIN

Query Match      100.0%; Score 21; DB 10; Length 139;
Best Local Similarity 100.0%; Pred. No. 6.2e+04;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 agaaacaacaacaacaaca 21
Db 14 AGAAACAACAACAACAACA 34

RESULT 4
BASE COUNT  54 a          29 c          20 g          20 t
ORIGIN

Query Match      100.0%; Score 21; DB 10; Length 100;

```

LOCUS B1043282 144 bp mRNA linear EST 14-JUN-2001  
 DEFINITION PM4-OT0201-120301-001-003\_1 OT0201 Homo sapiens cDNA, mRNA  
 sequence.  
 ACCESSION B1043282  
 VERSION B1043282.1 GI:14450012  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 144)  
 Dias Neto, E., Garcia Correa, R., Veijovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 CONTACT: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM4&t2=PM4-OT0201-  
 120301-001-d03-1&t3=2001-03-12&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 138.  
 Location/Qualifiers  
 1..144  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="OT0201"  
 /dev\_stage="Adult"  
 /note="Organ: ovary; Vector: puc18; Site\_1: SmaI; Site\_2:  
 SmaI; A mini-library was made by cloning products derived  
 from ORESTES PCR (U.S. Letters Patent application No. 196  
 716 - Ludwig Institute for Cancer Research) profiles  
 into the pUC 18 vector. Reverse transcription of tissue  
 mRNA and cDNA amplification were performed under low  
 stringency conditions."  
 35 c 30 g 26 t  
 BASE COUNT 53 a 35 c 30 g 26 t  
 ORIGIN

Query Match 100.0%; Score 21; DB 10; Length 144;  
 Best Local Similarity 100.0%; Pred. No. 6e+04;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaacaaca 21  
 ||||||||||||||||||||  
 Db 56 AGAACAACAACAACAACA 76

RESULT 5  
 LOCUS A1345073 148 bp mRNA linear EST 08-APR-1999  
 DEFINITION t062b07.x1 NCI-CGAP\_Brn20 Homo sapiens cDNA clone IMAGE:2058901 3',  
 mRNA sequence.  
 ACCESSION A1345073  
 VERSION A1345073.1 GI:4082279  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 148)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: cgaps-r@mail.nih.gov  
 Tissue Procurement: Robert Jenkins, M.D., Mark Israel, M.D., Jim  
 Jacobson, Ph.D.  
 cDNA Library Preparation: David B. Krizman, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ILN at:  
 www-bio.lnl.gov/bdrp/image/image.html  
 Insert Length: 237 Std Error: 0.00  
 Seq primer: -400P from gibco.  
 Location/Qualifiers  
 1..148  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="IMAGE:2058901"  
 /clone\_lib="NCI-CGAP\_Brn20"  
 /tissue\_type="Oligodendroglioma"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pAMP1; mRNA made from  
 Oligodendroglioma tissue, cDNA made by oligo-dt priming.  
 Directionally cloned. Size selected on agarose gel,  
 average insert size 500 bp. Primary library,  
 non-amplified."  
 26 c 30 g 29 t  
 BASE COUNT 63 a 26 c 30 g 29 t  
 ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 6e+04;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaacaaca 21  
 ||||||||||||||||||||  
 Db 52 AGAACAACAACAACAACA 72

RESULT 6  
 LOCUS BF922085 149 bp mRNA linear EST 19-JAN-2001  
 DEFINITION CM2-NT0170-181100-536-e05 NT0170 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BF922085  
 VERSION BF922085.1 GI:12317973  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 149)  
 Dias Neto, E., Garcia Correa, R., Veijovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 CONTACT: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001

Email: asimpson@ludwig.org br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&tl=CM2-NT0170-  
 181100-536-e05&tl=2000-11-18&tl=1)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 149.

## FEATURES

## SOURCE

1. 149  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NT0170"  
 /dev\_stage="Adult"

/note="Organ: nervous tumor; Vector: puc18; Site: 1; SmaI;  
 Site: 2; SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."

BASE COUNT 35 a 39 c 32 g 43 t  
 ORIGIN

Query Match 100.0%; Score 21; DB 10; Length 149;  
 Best Local Similarity 100.0%; Pred. No. 5.9e+04;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 aagaacaacaacaacaaca 21  
 ||||||||||||||||||||  
 Db 62 AGAACAACAACAACAACA 42

RESULT 7 150 bp mRNA EST 23-OCT-1997  
 AA639520 n991a09.81 NCI\_CGAP\_C09 Homo sapiens cDNA clone IMAGE:1159672 3',  
 LOCUS mRNA sequence.  
 ACCESSION AA639520  
 VERSION AA639520.1 GI:2563299  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 150)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 Tumor Gene Index  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Unpublished (1997)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: cgaaps-remail.nih.gov  
 Tissue Procurement: Ilian Kirsch, M.D., Michael R. Emmert-Buck, M.D.,  
 Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www.bio.llnl.gov/bdip/image/image.html  
 Seq primer: -40ml3 fwd. ET from Amersham.

## FEATURES

## SOURCE

1. 150  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1159672"  
 /clone\_lib="NCI\_CGAP\_C09"  
 /tissue\_type="colon tumor RER+"  
 /lab\_host="DH10B"  
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a  
 modified polylinker; 1st strand cDNA was prepared from  
 RER+ colon tumor, and was then primed with a Not I -  
 oligo(CT) primer. Double-stranded cDNA was ligated to Eco

BASE COUNT 29 a 18 c 53 g 50 t  
 ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 5.9e+04;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 aagaacaacaacaacaaca 21  
 ||||||||||||||||||||  
 Db 111 AGAACAACAACAACAACA 91

RESULT 8 156 bp DNA GSS 05-OCT-2000  
 AZ493323 1M0327M24R Mouse 10kb plasmid UUC1M library Mus musculus genomic  
 LOCUS clone UUC1M0327M24 R. DNA sequence.  
 DEFINITION  
 ACCESSION AZ493323.1 GI:10666890  
 VERSION  
 KEYWORDS  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sclerogasthi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 156)  
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamli, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, J., Reilly,  
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.  
 and Wright, D., Weiss, R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 Unpublished (2000)  
 JOURNAL Contact: Robert B. Weiss  
 COMMENT University of Utah  
 Bm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT  
 84112, USA  
 Tel: 801 385 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00  
 Plate: 0327 row: M column: 24  
 Seq primer: CACACAGCAACACATATGACC  
 Class: Plasmid ends  
 High quality sequence stop: 156.

## FEATURES

## SOURCE

1. 156  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10990"  
 /clone="UUC1M0327M24"  
 /clone\_lib="Mouse 10kb plasmid UUC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: pMD42ny: Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pMD42 (g14732114|g14732072.1), a copy-number

Inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT 28 a 41 c 29 g 58 t  
ORIGIN

Query Match 100.0%; Score 21; DB 12; Length 156;  
Best Local Similarity 100.0%; Pred. No. 5.8e+04;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaacaacaacaaca 21  
|||||  
DB 113 AGAACAACAACAACAACA 93

## RESULT 9

LOCUS BH105347 156 bp DNA linear GSS 19-JUL-2001  
DEFINITION RPCI-24-379D17.TV RPCI-24 Mus musculus genomic clone RPCI-24-379D17  
, DNA sequence.

ACCESSION BH105347  
VERSION BH105347.1 GI:14934204  
KEYWORDS GSS.

## SOURCE

ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 156)  
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akimel, B., Levins, M.,  
Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,  
Russell, D., de Jong, P. and Fraser, C.M.

Mouse BAC End Sequences from library RPCI-24  
Unpublished (1999)  
Other GSSs: RPCI-24-379D17.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

## TITLE

JOURNAL

COMMENT

Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-24. For BAC  
library availability, please contact Pieter de Jong  
(pdejong@tigr.org). Clones may be purchased from BACPAC  
Resources (<http://www.tigr.org/bacpac/orderingframe.htm>). BAC end  
page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 379 row: D column: 17  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source Location/Qualifiers

1..156  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-24-379D17"  
/clone\_1ib="RPCI-24"  
/sex="Male"  
/cell\_type="Spleen/Brain"  
/note="Vector: pTARBAC1; Site\_1: BamHI; Site\_2: BamHI;  
RPCI-24 Mouse BAC library produced by Pieter de Jong. The  
library was cloned in the pTARBAC1 cloning vector at the  
BamHI sites using MboI partially digested male C57BL/6J  
DNA."

BASE COUNT 94 a 13 c 45 g 4 t  
ORIGIN

Query Match 100.0%; Score 21; DB 12; Length 156;  
Best Local Similarity 100.0%; Pred. No. 5.8e+04;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaacaacaacaaca 21  
|||||  
DB 30 AGAACAACAACAACAACA 50

## RESULT 10

LOCUS AI253971 167 bp mRNA linear EST 12-NOV-1998  
DEFINITION qv54a05.x1 NCI\_CGAP\_OV32 Homo sapiens CDNA clone IMAGE:1985360 3',  
mRNA sequence.

ACCESSION AI253971  
VERSION AI253971.1 GI:3861496  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 167)  
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps@mail.nih.gov  
Tissue procurement: W. Maistron Linehan, M.D., Rodrigo F. Chuang,  
M.D., Michael R. Emmert-Buck, M.D., Ph.D.

CDNA library preparation: David B. Kitzman, Ph.D.  
DNA sequencing by: The I.M.A.G.E. Consortium, LNL  
DNA distribution: NCI-CCGAP clone distribution information can be  
found through the I.M.A.G.E. consortium/LNL at:  
[www.bio.lnl.gov/bnlp/image/image.html](http://www.bio.lnl.gov/bnlp/image/image.html)  
Seq primer: 400P from Glibco.

FEATURES  
source Location/Qualifiers

1..167  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1985360"  
/clone\_1ib="NCI\_CGAP\_OV32"  
/sex="female"  
/tissue\_type="Papillary serous carcinoma"  
/lab\_host="DH10B"  
/note="Organ: ovary; Vector: pAMP1; mRNA made from ovarian  
carcinoma, CDNA made by Oligo-dT priming.  
Non-directionally cloned. Size-selected on agarose gel,  
average insert size 500 bp. Non-amplified library."  
BASE COUNT 79 a 23 c 31 g 34 t  
ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 167;  
Best Local Similarity 100.0%; Pred. No. 5.6e+04;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaacaacaacaaca 21  
|||||  
DB 52 AGAACAACAACAACAACA 72

## RESULT 11

LOCUS AI733948 167 bp mRNA linear EST 14-JUN-1999  
DEFINITION qv54a05.x5 NCI\_CGAP\_OV32 Homo sapiens CDNA clone IMAGE:1985360 3',  
mRNA sequence.

ACCESSION AI733948  
VERSION AI733948.1 GI:5055061  
KEYWORDS EST.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 167)  
 AUTHORS NCI/NIH-CCAP http://www.ncbi.nlm.nih.gov/nciccap.  
 TITLE National Cancer Institute / National Institute of Dental Research,  
 Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgrabs-r@mail.nih.gov  
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuquai,  
 M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: David B. Krizman, Ph.D.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium, LNL  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CCAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 www.bio.lnlnl.gov/bbrp/image/image.html

FEATURES  
 source  
 This read is a RESEQUENCE of a previously sequenced human clone  
 Original clone citation: National Cancer Institute, Cancer Genome  
 Anatomy Project (CGAP), Tumor Gene Index  
 This read has been verified (found to hit its original self in the  
 correct orientation)  
 Seq primer: -400P from Gibco.  
 Location/Qualifiers  
 1. 167  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1985360"  
 /clone\_lib="NCI-CCAP\_Ov32"  
 /sex="female"  
 /tissue\_type="papillary serous carcinoma"  
 /lab\_host="DH10B"  
 /note="Organ: ovary; Vector: PAMP1; mRNA made from ovarian  
 carcinoma, CDNA made by oligo-dT priming.  
 Non-directionally cloned. Size-selected on agarose gel,  
 average insert size 500 bp. Non-amplified library."

BASE COUNT 79 a 23 c 31 g 34 t  
 ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 167;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+04;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaaacaaacaaacaa 21  
 ||||||||||||||||||||  
 Db 52 AGAACAACAAACAAACAA 72

RESULT 12 177 bp mRNA linear EST 30-DEC-1997  
 AA719825  
 LOCUS zhs9d03.s1 Soares\_Pineal.gland.N3HPG Homo sapiens CDNA clone  
 DEFINITION IMAGE:414435 3' similar to contains Alu repetitive element; contains  
 element ITR1 repetitive element; mRNA sequence.

ACCESSION AA719825  
 VERSION AA719825.1 GI:2732924  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 177)  
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin,  
 J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,  
 White, Y., Wylie, T., Waterston, R. and Wilson, R.

TITLE WashU-NCI human EST Project  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: est@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnlnl.gov) for further information.  
 Seq primer: -40m13 fwd. Ex from Amersham  
 High quality sequence stop: 165.

FEATURES  
 source  
 Location/Qualifiers  
 1. 177  
 /organism="Homo sapiens"  
 /db\_xref="GDB:1319706"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:414435"  
 /clone\_lib="Soares\_Pineal.gland.N3HPG"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: pineal gland; Vector: p77T3D (Pharmacia)  
 with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
 1st strand cDNA was primed with a Not I - oligo(dT) primer  
 [5' TGTTACCAATCTGAAGTGCAGCGCCGCCGCTTTTCTTTTCTTTT 3']  
 double-stranded cDNA was size selected, ligated to Eco  
 RI adapters (Pharmacia), digested with Not I and cloned  
 into the Not I and Eco RI sites of a modified p77T3 vector  
 (Pharmacia). Library constructed by Bento Soares and  
 M. Patricia Bonaldo."

BASE COUNT 55 a 48 c 46 g 28 t  
 ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 177;  
 Best Local Similarity 100.0%; Pred. No. 5.4e+04;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaaacaaacaaacaa 21  
 ||||||||||||||||||||  
 Db 38 AGAACAACAAACAAACAA 58

RESULT 13 180 bp mRNA linear EST 06-JUN-1997  
 AA457363/c  
 LOCUS aa91f12.r1 StrataGene fetal retina 937202 Homo sapiens CDNA clone  
 DEFINITION IMAGE:838703 5' similar to contains element MRR22 repetitive  
 element; mRNA sequence.

ACCESSION AA457363  
 VERSION AA457363.1 GI:2180083  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 180)  
 AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Eukaryota; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
 Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,  
 Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie,  
 T., Waterston, R., and Wilson, R.

TITLE WashU-Merck EST Project 1997  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnlnl.gov) for further information.  
 Seq primer: -28m13 rev. Ex from Amersham  
 High quality sequence stop: 167.  
 Location/Qualifiers  
 1. 180  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:838703"  
 /clone\_lib="StrataGene fetal retina 937202"  
 /sex="mixed"  
 /lab\_host="SOLR (kanamycin resistant)"

/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Pooled retinal tissue. Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5' CAAATCGCAGACAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."

BASE COUNT 42 a 48 c 34 g 56 t

ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 180;  
Best Local Similarity 100.0%; Pred. No. 5.4e+04;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaacaacaacaacaaca 21  
|||||  
Db 171 AGAACAACAACAACAACA 151

# RESULT 14

LOCUS A1733998 181 bp mRNA linear EST 13-DEC-1999  
DEFINITION qv55g95.x5 NCI-CGAP\_OV32 Homo sapiens CDNA clone IMAGE:1985528 3',  
mRNA sequence.

ACCESSION A1733998  
VERSION A1733998.1 GI:5055111  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 181)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuagui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium, Ph.D.  
DNA sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www.bio.llnl.gov/bdrrp/image/image.html](http://www.bio.llnl.gov/bdrrp/image/image.html)

This read is a RESEQUENCE of a previously sequenced human clone  
Original clone citation: National Cancer Institute, Cancer Genome  
Anatomy Project (CGAP), Tumor Gene Index  
This read has been verified (found to hit its original self in the  
insert orientation)  
Insert length: 238 Std Error: 0.00  
Seq primer: -400P from GIBCO.

## FEATURES

source

Location/Qualifiers  
1. 181  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1985528"  
/clone\_1lb="NCI-CGAP\_OV32"  
/sex="Female"  
/tissue\_type="papillary serous carcinoma"  
/lab\_host="DH10B"  
/note="Organ: ovary; Vector: PAMPI; mRNA made from ovarian carcinoma, CDNA made by oligo-dT priming.  
Non-directionally cloned. Size-selected on agarose gel,  
average insert size 500 bp. Non-amplified library."  
BASE COUNT 78 a 32 c 34 g 37 t  
ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 181;  
Best Local Similarity 100.0%; Pred. No. 5.4e+04;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaacaacaacaacaaca 21  
|||||  
Db 75 AGAACAACAACAACAACA 95

# RESULT 15

LOCUS BB072882/c 184 bp mRNA linear EST 27-JUN-2000  
DEFINITION BB072882 RIKEN full-length enriched, adult male colon Mus musculus  
CDNA clone 9030619B06 3' similar to U43144 Mus musculus  
phospholipase C beta3 mRNA, mRNA sequence.

ACCESSION BB072882  
VERSION BB072882.1 GI:8582880  
KEYWORDS EST.

SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 184)

Kono H., Aizawa K., Akahira S., Akiyama J., Arakawa T., Carninci P., Endo T., Fukuda S., Fukunishi Y., Hara A., Hayatsu N., Hirozane T., Horii F., Ishii Y., Ishikawa J., Ishikawa T., Itoh M., Izawa M., Kadoya K., Kagawa I., Kai C., Kawai J., Kikuchi N., Kiyosawa H., Kojima Y., Kondo S., Koya S., Kurihara C., Kusakabe M., Matsuyama T., Miki R., Mizuno Y., Nakamura M., Oda H., Okazaki Y., Ono T., Owa C., Saito H., Sakai C., Satoh K., Shibata K., Shibata Y., Shigemoto Y., Shinagawa A., Shitaki T., Sogabe Y., Sugahara Y., Suzuki H., Suzuki H., Tagawa A., Takahashi F., Tomiwa N., Toya T., Tsunoda Y., Watabiki A., Watanabe S., Yamamura T., Yamanaka I., Yano R., Yasunishi A., Yokota T., Yoshida K., Yoshiki A., Yoshino M., Muramatsu M. and Hayashizaki Y.  
RIKEN Mouse ESTs (Kono H., et al.)  
Unpublished (2000)  
Contact: Yoshinori Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-resgsc.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>  
Carninci P., Nishiyama Y., Westover A., Itoh M., Nagaoka S., Sasaki N., Okazaki Y., Muramatsu M. and Hayashizaki Y.  
Thermotabilization and thermoinactivation of thermostable enzymes by CDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh M., Katsunai T., Akiyama J., Shibata K., Izawa M., Kawai J., Tomaru Y., Carninci P., Shibata Y., Ozawa Y., Muramatsu M., Okazaki Y. and Hayashizaki Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci P. and Hayashizaki Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

## FEATURES

source

Location/Qualifiers  
1. 184  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="9030619B06"  
/clone\_1lb="RIKEN full-length enriched, adult male colon"  
/sex="male"  
/tissue\_type="colon"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in

RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTCTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTTAATTAAATTAAATCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pluescript KS(+) after bulk excision from lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 43 a 37 c 45 g 59 t  
ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 184;  
Best Local Similarity 100.0%; Pred. No. 5.3e+04;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaacacaacaacaacaacaa 21  
|||||  
Db 154 AGAACCAACCAACCAACAA 134

Search completed: July 31, 2002, 12:11:57  
Job time: 13470 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:11:09 ; Search time 720.8 Seconds  
(without alignments)  
50.021 Million cell updates/sec

Title: US-09-899-718A-3

Perfect score: 21  
Sequence: 1 agaacacaacaacaacaa 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_032802:\*

- 1: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT.\*
- 2: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT.\*
- 3: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT.\*
- 4: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT.\*
- 5: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT.\*
- 6: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT.\*
- 7: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT.\*
- 8: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT.\*
- 9: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT.\*
- 10: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT.\*
- 11: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT.\*
- 12: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT.\*
- 13: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT.\*
- 14: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT.\*
- 15: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT.\*
- 16: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT.\*
- 17: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT.\*
- 18: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT.\*
- 19: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT.\*
- 20: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT.\*
- 21: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.\*
- 22: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.\*
- 23: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.\*
- 24: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| C 1        | 21    | 100.0       | 35     | 18 | AAT93839    |
| 2          | 21    | 100.0       | 140    | 21 | AACT13041   |
| 3          | 21    | 100.0       | 384    | 20 | AAV87203    |
| 4          | 21    | 100.0       | 770    | 21 | AAC44102    |
| 5          | 21    | 100.0       | 1106   | 22 | AAS46192    |
| 6          | 21    | 100.0       | 1122   | 22 | AAD09952    |
| C 7        | 21    | 100.0       | 1409   | 22 | ABA16382    |
| 8          | 21    | 100.0       | 1688   | 21 | AAC79000    |
| C 9        | 21    | 100.0       | 2224   | 16 | AA089544    |

|      |    |       |       |    |          |
|------|----|-------|-------|----|----------|
| C 10 | 21 | 100.0 | 2339  | 12 | AAQ10956 |
| C 11 | 21 | 100.0 | 2339  | 20 | AAZ09171 |
| C 12 | 21 | 100.0 | 2339  | 22 | AAH48860 |
| C 13 | 21 | 100.0 | 2357  | 22 | AAK85884 |
| C 14 | 21 | 100.0 | 2358  | 22 | AAK85885 |
| C 15 | 21 | 100.0 | 2358  | 22 | AAK85886 |
| C 16 | 21 | 100.0 | 2393  | 12 | AAQ10907 |
| C 17 | 21 | 100.0 | 2394  | 22 | AAK83951 |
| C 18 | 21 | 100.0 | 2592  | 22 | AAH15452 |
| C 19 | 21 | 100.0 | 2613  | 21 | AAH49207 |
| C 20 | 21 | 100.0 | 3001  | 21 | AAH51796 |
| C 21 | 21 | 100.0 | 3004  | 23 | ABL26224 |
| C 22 | 21 | 100.0 | 4265  | 22 | AAK53361 |
| C 23 | 21 | 100.0 | 4295  | 22 | AAI58456 |
| C 24 | 21 | 100.0 | 4379  | 22 | AAK53345 |
| C 25 | 21 | 100.0 | 4379  | 22 | AAI60242 |
| C 26 | 21 | 100.0 | 4676  | 23 | AAK84690 |
| C 27 | 21 | 100.0 | 4809  | 22 | AAK51993 |
| C 28 | 21 | 100.0 | 5532  | 22 | AAI04707 |
| C 29 | 21 | 100.0 | 6089  | 24 | AB199827 |
| C 30 | 21 | 100.0 | 7302  | 22 | AAK74467 |
| C 31 | 21 | 100.0 | 7815  | 22 | AAK89492 |
| C 32 | 21 | 100.0 | 9328  | 22 | AAK74469 |
| C 33 | 21 | 100.0 | 10679 | 22 | AAK89493 |
| C 34 | 21 | 100.0 | 10679 | 22 | AAK89493 |
| C 35 | 21 | 100.0 | 11811 | 20 | AAV83943 |
| C 36 | 21 | 100.0 | 16281 | 22 | AAK70314 |
| C 37 | 21 | 100.0 | 16285 | 22 | AAK70315 |
| C 38 | 21 | 100.0 | 16428 | 22 | AAK70316 |
| C 39 | 21 | 100.0 | 17679 | 22 | AAK81010 |
| C 40 | 21 | 100.0 | 17679 | 22 | AAI67271 |
| C 41 | 21 | 100.0 | 19472 | 22 | AAI67272 |
| C 42 | 21 | 100.0 | 21458 | 22 | AAI67274 |
| C 43 | 21 | 100.0 | 21458 | 22 | AAI67122 |
| C 44 | 21 | 100.0 | 31834 | 22 | AAK81266 |
| C 45 | 21 | 100.0 | 31952 | 22 | AAK89370 |

#### ALIGNMENTS

RESULT 1  
AAT93839/C  
ID AAT93839 standard; DNA; 35 BP.  
XX  
AC AAT93839;  
XX  
DT 24-FEB-1998 (first entry)  
XX  
DE Phosphodiester oligonucleotide 29 with cytotoxic activity.  
XX  
DE Phosphodiester; selective binding; cell viability; growth;  
XX  
KW tumoural cell line; cytotoxic activity; tumour cell; lymphoma;  
KW Lymphoblastic tumour; ss.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT modified\_base 1..35 /\*tag= a  
FT /\*note= "phosphodiester oligonucleotide"  
PN W09720924-A1.  
PD 12-JUN-1997.  
XX  
PF 04-DEC-1996; 96WO-EP05388.  
XX  
PR 04-DEC-1995; 95IT-OM12539.  
XX  
PA (SAIC-) SAICOM SRL.  
XX  
PI Quadrifoglio F, Scaggiante B;

XX DR WPI; 1997-319771/29.

XX PT New phospho-di-esteric oligo-nucleotide(s) - which exert a specific  
XX PT and selective cytotoxic effect on tumour cells, for treating both  
XX PT solid and liquid tumours

PS Example 5; Page 12; 38pp; English.

XX CC Novel phosphodiesteric oligonucleotides AAT93837-41 are based on the  
CC generic formula, in the 3'-5' or 5'-3' direction:  
CC (Gara)'a'-(cbnb)'b'-(gcnc)'c'-(gafnd)'d'-(gete)'e'-(grff)'f'-'  
CC (G-9tg)'g'-'N', where:  
CC N and N' = T or G, equal or different from each other;  
CC x = 0-8, equal or different from each other;  
CC a, b, c, d, e, f, and g = 0-10, equal or different from each other;  
CC a', b', c', d', e', f', and g' = 0-30, equal or different from each  
CC other;  
CC a'', b'', c'', d'', e'', f'', and g'' = 1-16, equal or different from  
CC each other;  
CC The oligonucleotides (see also AAT93811-27) are believed to selectively  
CC bind and sequester some proteins which are essential to the viability  
CC and growth of tumoural cell lines. They have specific and selective  
CC cytotoxic activity against tumour cells, and can be used for treating  
CC tumours of the liquid type, in particular of lymphoblastic origin, and of  
CC the solid type, in particular lymphomas. Oligonucleotides AAT93834-36  
CC were created to determine the relevance of the specific features of the  
CC above generic formula, in determining cytotoxic activity. The present  
CC sequence is different from the generic sequence in that the sequence has  
CC flanking fragments containing C and T bases. The results showed that  
CC sequences having flanking fragments containing C and T bases do not exert  
CC any significant cytotoxic activities.

SQ Sequence 35 BP; 0 A; 4 C; 5 G; 26 T; 0 other;

Query Match 100.0%; Score 21; DB 18; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaacaacaacaacaaca 21  
|||||  
Db 31 AGAACAACAACAACAACA 11

RESULT 2

ID AAC13041 standard; cDNA; 140 BP.

XX AC AAC13041;

DT 06-OCT-2000 (first entry)

XX DE Human secreted protein 5' EST, SEQ ID NO: 17116.

XX KM Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX KM gene therapy; chromosome mapping; ss.

XX OS Homo sapiens.

XX PN EP1033401-A2.

XX PD 06-SEP-2000.

XX PF 21-FEB-2000; 2000EP-0200610.

XX PR 26-FEB-1999; 99US-0122487.

XX PA (GSEST) GENSET.

XX PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX DR WPI; 2000-500381/45.

XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX PS Claim 1; SEQ ID 17116; 71pp + CD-ROM; English.

XX CC The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.

SQ Sequence 140 BP; 47 A; 38 C; 38 G; 17 T; 0 other;

Query Match 100.0%; Score 21; DB 21; Length 140;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaacaacaacaacaaca 21  
|||||  
Db 109 agaacaacaacaacaaca 129

RESULT 3

ID AAV87203 standard; cDNA; 384 BP.

XX AC AAV87203;

DT 27-APR-1999 (first entry)

XX DE EST clone BM40.

XX KM Expressed sequence tag; secreted protein; haematopoiesis regulator;  
XX KM tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
XX KM chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;  
XX KM receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

XX OS Homo sapiens.

XX PN WO9845435-A2.

XX PD 15-OCT-1998.

XX PF 10-APR-1998; 98WO-US06954.

XX PR 10-APR-1997; 97US-0835913.

XX PA (GENY) GENETICS INST INC.

XX PI Agostino MJ, Jacobs K, Lavallic ER, McCoy JM, Merberg D;

XX PT Racine LA, Spaulding V, Treacy M;

XX DR WPI; 1999-070076/06.

XX PT New polynucleotides encoding human secreted proteins - derived from  
XX PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
XX PT ovary, pituitary, retina and colon cDNA libraries

XX PS Claim 1; Page 487; 633pp; English.

XX CC This sequence represents an expressed sequence tag (EST), and is a

CC polynucleotide of the invention. The polynucleotides of the invention are  
CC all secreted EST sequences isolated from a variety of human tissue  
CC sources. The EST sequences and proteins encoded by them are predicted to  
CC have useful biological activities which would make them suitable for  
CC treating, preventing or ameliorating medical conditions in humans and  
CC animals, although no supporting data is given. Suggested activities  
CC include nutritional activity, immune stimulating or suppressing activity,  
CC haematopoiesis regulating activity, tissue growth or suppressing activity,  
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
CC therapy. The EST sequences are also stated to be useful for gene  
CC  
XX

SQ Sequence 384 BP; 89 A; 92 C; 80 G; 123 T; 0 other;

Query Match 100.0%; Score 21; DB 20; Length 384;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aagaacaacaacaacaaca 21  
|||||

Db 216 aagaacaacaacaacaaca 236

## RESULT 4

AAC44102 AAC44102 standard; DNA; 770 BP.

AC AAC44102;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 41638.

XX Hybridisation assay; genetic mapping; gene expression control;

KM protein identification; signal transduction pathway;

XX metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

PN EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0127462.

PR 01-APR-1999; 99US-0128234.

PR 06-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.

```

PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 20-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151303.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 10-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154779.
PR 20-SEP-1999; 99US-0155139.
PR 22-SEP-1999; 99US-0155486.
PR 23-SEP-1999; 99US-0156559.
PR 24-SEP-1999; 99US-0156558.
PR 28-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.

```

```

PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 21; DB 21; Length 770;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 agaacaacaacaacaaca 21
Db 11 agaacaacaacaacaaca 31

RESULT 5
AAS46192
ID AAS46192 standard; cDNA; 1106 BP.
XX AC
XX AAS46192;
XX DE
XX 18-DEC-2001 (first entry)
XX DE
XX Human DNA encoding PRO polypeptide sequence #268.
KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
KW PCR primer.
XX KW
XX Homo sapiens.
XX OS
XX WO200168848-A2.
XX PN
XX 20-SEP-2001.
XX PD
XX 28-FEB-2001; 2001WO-US06520.
XX PF
XX 01-MAR-2000; 2000WO-US05601.
XX PR 02-MAR-2000; 2000WO-US05841.
XX PR 03-MAR-2000; 2000US-187202P.
XX PR 06-MAR-2000; 2000US-186968P.
XX PR 14-MAR-2000; 2000US-189320P.
XX PR 14-MAR-2000; 2000US-189328P.
XX PR 15-MAR-2000; 2000WO-US06884.
XX PR 21-MAR-2000; 2000US-190828P.
XX PR 21-MAR-2000; 2000US-191007P.
XX PR 21-MAR-2000; 2000US-191048P.
XX PR 21-MAR-2000; 2000US-191314P.
XX PR 28-MAR-2000; 2000US-192655P.
XX PR 29-MAR-2000; 2000US-193032P.
XX PR 29-MAR-2000; 2000US-193053P.
XX PR 30-MAR-2000; 2000WO-US08439.
XX PR 04-APR-2000; 2000US-194449P.
XX PR 04-APR-2000; 2000US-194647P.
XX PR 11-APR-2000; 2000US-195975P.
XX PR 11-APR-2000; 2000US-196000P.
XX PR 11-APR-2000; 2000US-196187P.
XX PR 11-APR-2000; 2000US-196690P.
XX PR 11-APR-2000; 2000US-196620P.
XX PR 11-APR-2000; 2000US-198121P.
XX PR 18-APR-2000; 2000US-198585P.
XX PR 25-APR-2000; 2000US-199397P.
XX PR 25-APR-2000; 2000US-199550P.
XX PR 25-APR-2000; 2000US-199554P.
XX PR 03-MAY-2000; 2000US-201516P.
XX PR 17-MAY-2000; 2000WO-US13705.
XX PR 22-MAY-2000; 2000WO-US14042.

```

PR 30-MAY-2000; 2000MO-US14941.  
 PR 02-JUN-2000; 2000MO-US15264.  
 PR 05-JUN-2000; 2000US-209832P.  
 PR 28-JUL-2000; 2000MO-US20710.  
 PR 22-AUG-2000; 2000US-0644848.  
 PR 24-AUG-2000; 2000MO-US23328.  
 PR 08-NOV-2000; 2000MO-US30952.  
 PR 01-DEC-2000; 2000MO-US32678.  
 PR 20-DEC-2000; 2000MO-US34956.  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;  
 PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;  
 XX  
 DR WPI: 2001-602746/68.  
 DR P-PSDB; AAU29291.  
 XX  
 PT Novel nucleic acids encoding PRO polypeptides, used to diagnose the  
 PT presence of tumours, such as prostate and breast tumours, in mammals and  
 PT to screen for modulators of the compounds -  
 XX  
 PS Claim 2; Fig 535; 774pp; English.  
 XX  
 CC Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR  
 CC primers for PRO polypeptides of the invention. The sequences of the  
 CC invention can be used to detect the presence of a tumour in a mammal by  
 CC comparing the level of expression of a PRO polypeptide in a test sample  
 CC of cells from the animal and a control sample of normal cells, whereby a  
 CC higher level of expression in the test sample indicates the presence of a  
 CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,  
 CC pigs, goats and rabbits but are preferably human. The polypeptides can be  
 CC used to stimulate tumour necrosis factor (TNF) alpha release from human  
 CC blood, when contacted with it. A specific polypeptide can be used to  
 CC stimulate the proliferation or differentiation of chondrocyte cells. The  
 CC PRO proteins can be used to determine the presence of tumours and also  
 CC susceptibility to tumour development, particularly adrenal, lung, colon,  
 CC breast, prostate, rectal, cervical, or liver tumours. In mammalian  
 CC subjects the oligonucleotide probes specific for the PRO nucleic acids  
 CC can be used for genetic analysis of individuals with genetic disorders.  
 CC  
 XX Sequence 1106 BP; 325 A; 226 C; 260 G; 295 T; 0 other;  
 SQ  
 Query Match 100.0%; Score 21; DB 22; Length 1106;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 agaacaacaacaacaaca 21  
 ||||||||||||||||||  
 Db 845 agaacaacaacaacaaca 865  
 RESULT 6  
 AAD09952  
 ID AAD09952 standard; CDNA; 1122 BP.  
 XX  
 AC AAD09952;  
 XX  
 DT 12-SEP-2001 (first entry)  
 DE Human drug metabolising enzyme (DME-17) cDNA.  
 XX  
 KW Human; drug metabolising enzyme; DME-17; immunosuppressive; gene therapy;  
 KW cytosolic; autoimmune disorder; inflammatory disorder; atherosclerosis;  
 KW osteoporosis; eye disorder; hepatic tumour; Addison's disease; cretinism;  
 KW rheumatoid arthritis; acquired immune deficiency syndrome; AIDS; anaemia;  
 KW developmental disorder; endocrine disorder; iritis; acromegaly; epilepsy;  
 KW thyrotoxicosis; pancreatic disorder; diabetes mellitus; obesity; adenoma;  
 KW gastrointestinal disorder; nodular hyperplasia; conjunctivitis; glaucoma;  
 KW actinic keratosis; metabolic disorder; dysphagia; anorexia; carcinoma;  
 XX cell proliferative disorder; ss.

OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 123..755  
 FT /\*tag= a  
 FT /product= "Human drug metabolising enzyme (DME-17)"  
 FT sig\_peptide 123..212  
 FT /\*tag= b  
 FT mat\_peptide 213..752  
 FT /\*tag= c  
 FT /product= "Mature drug metabolising enzyme (DME-17)"  
 XX  
 PN WO200151638-A2.  
 XX  
 PD 19-JUL-2001.  
 XX  
 PE 12-JAN-2001; 2001MO-US01174.  
 XX  
 PF 14-JAN-2000; 2000US-0176139.  
 PR 21-JAN-2000; 2000US-0177443.  
 PR 28-JAN-2000; 2000US-0178574.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Yang J, Baughn MR, Burford N, Au-Young J, Lu DAM, Reddy R;  
 PI Ring HZ, Hillman JL, Yue H, Azimzal Y, Tao MG, Gandhi AR;  
 PI Nguyen DB, Tang YT, Lal P, Bandman O;  
 XX  
 DR WPI: 2001-425874/45.  
 DR P-PSDB; AAE05186.  
 XX  
 PT Drug metabolizing enzymes and encoding polynucleotides, useful for  
 PT diagnosing, treating and/or preventing autoimmune, inflammatory, cell  
 PT proliferative, developmental, endocrine, eye, metabolic, and  
 PT gastrointestinal disorders -  
 XX  
 PS Claim 5; Page 168; 133pp; English.  
 XX  
 CC The present sequence is human drug metabolising enzyme (DME-17) cDNA.  
 CC Human DME and its nucleic acid molecule are useful for the diagnosis,  
 CC treatment and prevention of disorders associated with increased or  
 CC decreased expression of DME. Examples of such disorders include,  
 CC autoimmune/inflammatory disorder such as acquired immune deficiency  
 CC syndrome (AIDS), rheumatoid arthritis, osteoporosis; cell proliferative  
 CC disorder such as actinic keratosis, atherosclerosis; developmental  
 CC disorder such as epilepsy, anaemia; endocrine disorder such as  
 CC acromegaly, cretinism, thyrotoxicosis; pancreatic disorder such as  
 CC diabetes mellitus; eye disorder such as conjunctivitis, glaucoma, iritis;  
 CC metabolic disorder such as Addison's disease, obesity; gastrointestinal  
 CC nodular hyperplasia, adenomas and hepatic tumours including  
 CC creating 'knockin' humanised animals (pigs) or transgenic animals (mice  
 CC or rats) to model human disease. DME DNA is also useful for gene  
 CC therapy. DME and its immunogenic fragments are useful for screening  
 CC libraries of compounds in several drug screening assays.  
 CC  
 XX Sequence 1122 BP; 338 A; 228 C; 261 G; 295 T; 0 other;  
 SQ  
 Query Match 100.0%; Score 21; DB 22; Length 1122;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 agaacaacaacaacaaca 21  
 ||||||||||||||||||  
 Db 845 agaacaacaacaacaaca 865  
 RESULT 7  
 ABA16382/c  
 ID ABA16382 standard; DNA; 1409 BP.  
 XX  
 AC ABA16382;

XX 23-JAN-2002 (first entry)  
XX Human nervous system related polynucleotide seq ID NO 6713.  
DE  
XX  
XX Human; nootropic; neuroprotective; cytosolic; dermatological; virucide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnereary;  
KW antiparkinsonian; antischizoid; antianemic; antitubercular; cancer;  
KW antineoplastic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antileuc; anticonvulsant; antifungal;  
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotoxic; gene therapy; vaccine; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200159063-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01334.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0186874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225759.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226888.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 20-OCT-2000; 2000US-0242221.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246612.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249224.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249246.  
PR 17-NOV-2000; 2000US-0249247.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.

PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251889.  
 PR 11-DEC-2000; 2000US-0251990.  
 PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-54165/60.

XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating nervous system  
 XX cancers and metastases -

XX Disclosure; SEQ ID NO 8713; 1701pp + Sequence Listing; English.

XX The invention relates to novel genes (AB11004-ABA21534) and proteins  
 CC (AB114678-AB18001) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1409 BP; 361 A; 334 C; 324 G; 390 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 1409;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaacaacaacaacaaca 21  
 ||||||||||||||||||||  
 DB 676 AGAACAACAACAACA 656

RESULT 8  
 AAC79000  
 ID AAC79000 standard; DNA; 1688 BP.

XX AAC79000;

XX 14-FEB-2001 (first entry)

XX Human secreted protein gene 4 clone HCBG349.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KM antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;  
 KM vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KM cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KM neurological disease; infection; human; secreted protein; ss.

XX Homo sapiens.

OS  
 XX

PN WO200058358-A1.

XX 05-OCT-2000.

XX 23-MAR-2000; 2000WO-US07725.

XX 26-MAR-1999; 99US-0126602.

XX 14-JAN-2000; 2000US-0176063.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsu G;

XX WPI; 2000-594640/56.

XX P-PSDB; AAB44338.

XX Fourty nine nucleic acid molecules encoding human secreted proteins,  
 PT useful in the prevention, treatment and diagnosis of cancer, immune  
 PT disorders, cardiovascular disorders and neurological diseases -

XX Claim 1; Page 317; 367pp; English.

XX The invention relates to the isolation of genes AAC78997-C79045 encoding  
 CC 49 human secreted proteins AAB44335-B44382. The genes can be used to  
 CC generate fusion proteins by linking to the gene for the human  
 CC immunoglobulin G Fc portion (AAC78988) for increasing the stability of  
 CC the fusion protein as compared to the human protein only. The genes and  
 CC proteins are useful for preventing, ameliorating or treating medical  
 CC conditions, e.g. by protein or gene therapy. The genes are isolated  
 CC from a range of human tissues disclosed in the specification. The  
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in  
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
 CC (d) wound healing; (e) neurological disorders such as myocardial ischaemias; (d)  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.

XX Sequence 1688 BP; 516 A; 310 C; 384 G; 478 T; 0 other;

Query Match 100.0%; Score 21; DB 21; Length 1688;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaacaacaacaacaaca 21  
 ||||||||||||||||||||  
 DB 875 agaacaacaacaacaaca 895

RESULT 9  
 AA089544/C  
 ID AA089544 standard; DNA; 2224 BP.

XX AA089544;

XX 31-OCT-1995 (first entry)

XX P75 Tumour Necrosis Factor Receptor.

XX Ligand; tumour necrosis factor; nerve growth factor; TNF; NGF;

XX receptor; ss..

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 90..1475

XX /tag=a

XX /product= p75 TNF receptor.

```

FT misc_difference 1137..1139
FT /tag= b
FT /transl_except= GCA encodes glycine.
FT misc_difference 1140..1142
FT /tag= c
FT /transl_except= CCA encodes Alanine.
FT misc_difference 1146..1148
FT /tag= d
FT /transl_except= GTG encodes Glutamic acid.
FT misc_difference 1149..1151
FT /tag= e
FT /transl_except= GAG encodes Alanine.
FT misc_difference 1152..1154
FT /tag= f
FT /transl_except= GCC encodes Arginine.
FT misc_difference 1155..1157
FT /tag= g
FT /transl_except= AGT encodes Alanine.
FT misc_difference 1158..1160
FT /tag= h
FT /transl_except= GGG encodes Serine.
FT misc_difference 1161..1163
FT /tag= i
FT /transl_except= GCC encodes Threonine.
FT misc_difference 1167..1169
FT /tag= j
FT /transl_except= GAG encodes Serine.
FT misc_difference 1170..1172
FT /tag= k
FT /transl_except= GCC encodes Serine.
FT misc_difference 1173..1175
FT /tag= l
FT /transl_except= CGG encodes Aspartic acid.
FT misc_difference 1176..1178
FT /tag= m
FT /transl_except= GCC encodes Serine.
FT misc_difference 1182..1184
FT /tag= n
FT /transl_except= ACC encodes Proline.
FT misc_difference 1188..1190
FT /tag= o
FT /transl_except= AGC encodes Glycine.
FT misc_difference 1191..1193
FT /tag= p
FT /transl_except= TCA encodes Histidine.
FT misc_difference 1194..1196
FT /tag= q
FT /transl_except= GAT encodes Glycine.
FT misc_difference 1197..1199
FT /tag= r
FT /transl_except= TCT encodes Threonine.
FT misc_difference 2000..2002
FT /tag= s
FT /transl_except= TCC encodes Glutamine.
FT misc_difference 2003..2005
FT /tag= t
FT /transl_except= CCT encodes Alanine.
FT misc_difference 2006..2008
FT /tag= u
FT /transl_except= GGT encodes Proline.
FT misc_difference 2012..2014
FT /tag= v
FT /transl_except= CAT encodes Valine.
FT misc_difference 2015..2016
FT /tag= w
FT /transl_except= GGG encodes Glutamic acid.
FT misc_difference 2017..2018
FT /tag= x
FT /transl_except= ACC encodes Alanine.
FT misc_difference 2019..2021
FT /tag= y
FT /transl_except= CAG encodes Serine.
XX

```

```

PN EP648783-A.
XX
XX PD 19-APR-1995.
XX PF 11-OCT-1994; 94EP-0116015.
XX PR 12-OCT-1993; 93II-0107267.
XX PA (YEDA ) YEDA RES & DEV CO LTD.
XX PA (WALL/) WALLACH D.
XX PI Beletsky I, Bigda J, Mett I, Wallach D;
XX WPI: 1995-148673/20.
XX DR P-PSDR; AAR72504.
XX PT Tumour necrosis factor (TNF) receptor ligand - used to increase
XX inhibitory effect of a soluble TNF receptor
XX PS Disclosure; Figure 2; 18pp; English.
XX CC A ligand to a member of the tumour necrosis factor (TNF)/nerve
XX CC growth factor (NGF) receptor family which binds either to the region
XX CC of the 4th-Cys rich domain of the receptor, or to the region between
XX CC it and the cell membrane may be used in the production of a
XX CC pharmaceutical composition for increasing the inhibitory effect of a
XX CC soluble receptor of the TNF/NGF receptor family. This sequence
XX CC encodes the p75 TNF receptor. N in the sequence represents an
XX CC unidentified nucleotide (poor reproduction in specification).
XX SQ Sequence 2224 BP; 432 A; 697 C; 688 G; 400 T; 7 other;

```

```

Query Match 100.0%; Score 21; DB 16; Length 2224;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaacaaca 21
DB 1847 AGAACAACAACAACAACA 1827

RESULT 10
AAQ10956/C
ID AAQ10956 standard; DNA; 2339 BP.
XX AC AAQ10956;
XX 24-MAY-1991 (first entry)
DE Encodes human 75kd TNF-binding protein.
XX XX
XX Tumour Necrosis Factor; binding proteins; septic shock;
KW autoimmune glomerulonephritis; lymphokine; cytokine.
FH Key 1..1179
FT CDS /tag= a
FT /product= 75kd TNF-BP
FT
XX EP417563-A.
XX 20-MAR-1991.
XX PD 31-AUG-1990; 90EP-0116707.
XX PF 20-APR-1990; 90CH-0001347.
XX PR 12-SEP-1989; 89CH-0003319.
XX PR 08-MAR-1990; 90CH-0000746.
XX XX
XX (HOFF ) HOFFMANN-LA ROCHE AG.
XX PA Brockhaus M, Dembic Z, Gentz R, Lesslauer W, Lotscher H;
XX

```



PI Schlaeger EJ;  
 XX WPI, 1991-081851/12.  
 DR P-PSDB; AAR11605.  
 XX  
 PT Insoluble tumor necrosis factor binding proteins - and DNA  
 PT encoding them, useful in pharmaceutical prods. and for antibody  
 PT prodt.  
 XX  
 PS Claim 4; Fig 1; 26pp; German.  
 XX  
 CC Partial amino acid sequences were determined for the 55 and 75kd  
 CC TNF-BPs (see AAR11072-R11081) and oligonucleotide primers were  
 CC synthesised based on these partial sequences. The primers were used  
 CC to produce a cDNA fragment for use as a probe to screen a human  
 CC placental cDNA bank constructed in lambda gt11. Positive clones were  
 CC identified and sequenced. Repeated sequencing showed a discrepancy  
 CC at position 7 such that the third codon encodes either Thr or Ser.  
 CC DNA constructs comprising the TNF-BP coding sequence may also  
 CC contain a fragment encoding a human Ig domain. Recombinant  
 CC constructs are used to transform cells to confer improved TNF-  
 CC binding properties.  
 CC See also AAQ10955.  
 CC  
 SQ Sequence 2339 BP; 494 A; 720 C; 685 G; 439 T; 1 other;

Query Match 100.0%; Score 21; DB 12; Length 2339;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaacaacaacaaca 21  
 ||||||||||||||||||||  
 Db 1550 AGAAACAAACAAACAA 1530

RESULT 11  
 AAZ09171/c  
 ID AAZ09171 standard; cDNA; 2339 BP.  
 AC AAZ09171;  
 DT 18-OCT-1999 (first entry)  
 DE Human tumour necrosis factor binding protein cDNA fragment.  
 KW Tumour necrosis factor binding protein; TNF; insoluble protein; agonist;  
 KW anti-inflammatory; antimalarial; treatment; septic shock; inflammation;  
 KW autoimmune glomerulonephritis; cerebral malaria; immune response;  
 KW antagonist; diagnosis; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT 1..1179  
 FT /\*tag= a  
 FT /product= "TNF binding protein"  
 FT /note= "Partial sequence, no start codon given"  
 XX  
 PN EP939121-A2.  
 PD 01-SEP-1999.  
 PF 31-AUG-1990; 90EP-0116707.  
 PR 20-APR-1990; 90CH-0001347.  
 PR 12-SEP-1989; 89CH-0003319.  
 PR 08-MAR-1990; 90CH-0000746.  
 XX  
 PA (HOFF) HOFFMANN LA ROCHE & CO AG F.  
 PI Brockhaus M, Dembic Z, Gentz R, Lesslauer W, Loetscher H;  
 PI Schlaeger E;

XX  
 DR WPI, 1999-480840/41.  
 DR P-PSDB; AAY30935.  
 XX  
 PT New insoluble proteins, and fragments, that bind to tumor necrosis  
 PT factor, used to treat e.g. septic shock or cerebral malaria  
 PT  
 PS Claim 4a; Fig 4; 25pp; German.  
 XX  
 CC This invention describes novel homogeneous insoluble proteins (I),  
 CC their (in)soluble fragments (Ia) and their salts that can bind tumor  
 CC necrosis factor (TNF). The products of the invention have  
 CC anti-inflammatory and antimalarial activity. (I) and (Ia) are used (i)  
 CC to treat diseases in which TNF is involved (e.g. septic shock, autoimmune  
 CC glomerulonephritis, cerebral malaria, immune responses and inflammation),  
 CC (ii) to purify TNF, (iii) to identify TNF (ant)agonists and (iv) for  
 CC diagnostic determination of TNF in body fluids. Antibodies raised against  
 CC (I) are used for affinity purification of (I). This sequence encodes  
 CC a tumor necrosis factor binding protein fragment described in the method  
 CC of the invention.  
 CC  
 SQ Sequence 2339 BP; 494 A; 720 C; 685 G; 440 T; 0 other;

Query Match 100.0%; Score 21; DB 20; Length 2339;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaacaacaacaaca 21  
 ||||||||||||||||||||  
 Db 1550 AGAAACAAACAAACAA 1530

RESULT 12  
 AAH48660/c  
 ID AAH48660 standard; DNA; 2339 BP.  
 AC AAH48660;  
 DT 12-NOV-2001 (first entry)  
 DE Human TNFBP-associated DNA #2.  
 DE  
 KW TNF; tumor necrosis factor binding protein; TNFBP; treatment;  
 KW insoluble protein; anti-inflammatory; immunosuppressive; antibacterial;  
 KW antiprotozoal; treatment; meningococcal sepsis; cerebral malaria;  
 KW autoimmune glomerulonephritis; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT 1..1179  
 FT /\*tag= a  
 FT /product= "TNFBP-associated protein"  
 XX  
 PN EP132471-A2.  
 PD 12-SEP-2001.  
 PF 31-AUG-1990; 2001EP-0108117.  
 PR 12-SEP-1989; 89CH-0003319.  
 PR 08-MAR-1990; 90CH-0000746.  
 PR 20-APR-1990; 90CH-0001347.  
 PR 31-AUG-1990; 90EP-0116707.  
 PR 31-AUG-1990; 99EP-0100703.  
 XX  
 PA (HOFF) HOFFMANN LA ROCHE & CO AG F.  
 PI Brockhaus M, Dembic Z, Gentz R, Lesslauer W, Loetscher H;  
 PI Schlaeger E;  
 DR WPI, 2001-559312/63.

DR P-PSDB; AAB86818.  
XX  
PR New homogeneous, insoluble proteins that bind tumor necrosis factor  
PR (TNF), useful for treating TNF-mediated disorders, e.g. inflammation  
XX  
PS Claim 4a; Fig 4; 26pp; German.  
XX  
CC This invention describes novel insoluble proteins (I), also their  
CC (in)soluble fragments and pharmaceutically acceptable salts, able to bind  
CC tumor necrosis factor (TNF) and in homogeneous form. The products of the  
CC invention have antiinflammatory, immunosuppressive, antibacterial,  
CC antiprotocoll activity. (I), and related recombinant proteins, are used  
CC to treat diseases mediated by TNF, e.g. shock in cases of meningococcal  
CC sepsis; development of autoimmune glomerulonephritis and cerebral  
CC malaria. Also (I), or antibodies specific for them, are used for  
CC diagnostic determination of TNF in body fluids, for affinity purification  
CC of TNF and for identifying (ant)agonists of TNF. This sequence encodes a  
CC human TNF binding protein described in the method of the invention.  
CC  
XX  
SQ Sequence 2339 BP; 494 A; 720 C; 685 G; 440 T; 0 other:  
  
Query Match 100.0%; Score 21; DB 22; Length 2339;  
Best Local Similarity 100.0%; Pred. No. 1; Le+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 agaacaacaacaacaaca 21  
|||||  
DB 1550 AGAACAACAACAACAACA 1530  
  
RESULT 13  
AAK85884/C  
ID AAK85884 standard; DNA; 2357 BP.  
XX  
AC AAK85884;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40696.  
XX  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01354.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216860.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229877.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234224.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0235484.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0238935.  
PR 13-OCT-2000; 2000US-0238937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246529.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249254.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250191.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating cancers and  
XX metastasis.  
XX  
XX Disclosure; SEQ ID NO 40696; 3071pp + Sequence Listing; English.  
XX  
XX AA54951 to AA64702 encode the human immune/hematopoietic antigen (I)  
XX amino acid sequences given in AA82170 to AA91921. (I) have cytostatic  
XX activity, and can be used in gene therapy and vaccine production. (I)  
XX proteins and polynucleotides may be used in the prevention, diagnosis and  
XX treatment of diseases associated with inappropriate (I) expression. For  
XX example, they may be used to treat disorders associated with decreased  
XX expression by rectifying mutations or deletions in a patient's genome  
XX that affect the activity of (I) by expressing inactive proteins or to  
XX supplement the patient's own production of (I). Additionally, (I)  
XX polynucleotides may be used to produce the secreted (I), by inserting  
XX the nucleic acids into a host cell and culturing the cell to express the  
XX protein. (I) proteins and polynucleotides may be used to prevent,  
XX cancer and treat immune/hematopoietic-related diseases, especially  
XX to AA87694 represent human immune/hematopoietic antigen genomic  
XX sequences from the present invention. AA54942 to AA54950 and AA82169  
XX represent sequences used in the exemplification of the present invention.

Sequence 2357 BP; 677 A; 465 C; 347 G; 868 T; 0 other;  
Query Match 100.0%; Score 21; DB 22; Length 2357;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 agaacaacaacaacaaca 21  
Db 427 AGAACACACACACACACAA 407  
RESULT 14  
AAK85885/C  
ID AAK85885 standard; DNA; 2358 BP.  
XX  
AC AAK85885;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human immune/hematopoietic antigen genomic sequence SEQ ID NO:40697.  
XX  
XX Human; immune; hematopoietic; immune/hematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX Homo sapiens.  
OS  
XX WO200157182-A2.  
PN  
XX 09-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US01354.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.



XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40698.  
DE  
XX Human: immune: haematopoietic; immune/haematopoietic antigen; cancer;  
KW Cytostatic; gene therapy; vaccine; metastasis; ds.  
XX Homo sapiens.  
OS  
XX WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01354.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239337.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249246.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.

```

PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
XX
XX PS Disclosure; SEQ ID NO 40698; 3071pp + Sequence Listing; English.
XX
XX CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX represent sequences used in the exemplification of the present invention.
XX
XX SQ Sequence 2358 BP; 679 A; 464 C; 347 G; 868 T; 0 other;

```

```

Query Match 100.0%; Score 21; DB 22; Length 2358;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 agaaacaacaacaacaaca 21
   |||||||||||||||||||
DB 427 AGAAACAACAACAACAACA 407

```

Search completed: July 31, 2002, 14:11:11  
 Job time: 17368 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:00:49 ; Search time 6034.22 Seconds  
(without alignments)  
72.828 Million cell updates/sec

Title: US-09-899-718A-3

Perfect score: 21

Sequence: 1 agaacaacaacaacaaca 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues 3595312

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:\*

- 1: gb\_da:\*
- 2: gb\_hcg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vl:\*
- 15: em\_da:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vl:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No. Query Match length DB ID Description

|    |    |       |      |    |            |                    |
|----|----|-------|------|----|------------|--------------------|
| 1  | 21 | 100.0 | 21   | 6  | AX349065   | AX349065 Sequence  |
| 2  | 21 | 100.0 | 116  | 10 | RNU76838   | U76838 Rattus norv |
| 3  | 21 | 100.0 | 120  | 10 | RNU76839   | U76839 Rattus norv |
| 4  | 21 | 100.0 | 224  | 8  | AY023679   | AY023679 Oryza sat |
| 5  | 21 | 100.0 | 239  | 11 | G35045     | G35045 STS NR03-14 |
| 6  | 21 | 100.0 | 281  | 11 | AU026129   | AU026129 Rattus no |
| 7  | 21 | 100.0 | 309  | 9  | HUMRPTPLM  | L20486 Human micro |
| 8  | 21 | 100.0 | 381  | 11 | HUMUT925   | L16357 Human chrom |
| 9  | 21 | 100.0 | 407  | 11 | HUMUT1400  | L17810 Human STS U |
| 10 | 21 | 100.0 | 497  | 9  | HUMDS8320T | L12265 Homo sapien |
| 11 | 21 | 100.0 | 497  | 11 | HUMUT5307  | L18590 Human chrom |
| 12 | 21 | 100.0 | 503  | 11 | G24380     | G24380 human STS W |
| 13 | 21 | 100.0 | 552  | 11 | G42246     | G42246 58P187 Hum  |
| 14 | 21 | 100.0 | 561  | 11 | HUMUT6478  | L30582 Human STS U |
| 15 | 21 | 100.0 | 617  | 5  | AF254660   | AF254660 Tetrao ur |
| 16 | 21 | 100.0 | 729  | 9  | HS297552   | AJ287552 Homo sapi |
| 17 | 21 | 100.0 | 766  | 8  | AY054247   | AY054247 Arabidops |
| 18 | 21 | 100.0 | 772  | 10 | MMU57395   | U57395 Mus musculu |
| 19 | 21 | 100.0 | 804  | 9  | HSNMF4GD01 | AF133493 Homo sapi |
| 20 | 21 | 100.0 | 870  | 11 | G15915     | G15915 human STS C |
| 21 | 21 | 100.0 | 994  | 10 | MMU91A5    | AF040950 Mus muscu |
| 22 | 21 | 100.0 | 1122 | 6  | AX195188   | AX195188 Sequence  |
| 23 | 21 | 100.0 | 1197 | 10 | AF199003S2 | AF199004 Mus muscu |
| 24 | 21 | 100.0 | 1213 | 10 | RATGSTPDS  | M14364 Rat glutath |
| 25 | 21 | 100.0 | 1687 | 5  | X1038844   | U38844 Xenopus lae |
| 26 | 21 | 100.0 | 1903 | 5  | GGU31223   | U31223 Gallus gall |
| 27 | 21 | 100.0 | 1970 | 10 | RATR2RB    | L14610 Rat transcr |
| 28 | 21 | 100.0 | 2084 | 5  | CCJUN      | X15547 Quail mRNA  |
| 29 | 21 | 100.0 | 2224 | 6  | AR152033   | AR152033 Sequence  |
| 30 | 21 | 100.0 | 2253 | 6  | A78517     | A78517 Sequence 1  |
| 31 | 21 | 100.0 | 2339 | 6  | A26415     | A26415 CDNA fragme |
| 32 | 21 | 100.0 | 2390 | 10 | MMMH07G1   | X03210 Mouse MHC Q |
| 33 | 21 | 100.0 | 2394 | 9  | HUMTNPRII  | M55994 Human tumor |
| 34 | 21 | 100.0 | 2592 | 9  | AK021890   | AK021890 Homo sapi |
| 35 | 21 | 100.0 | 2613 | 9  | HSTNER2S10 | U52165 Human tumor |
| 36 | 21 | 100.0 | 2831 | 8  | PBL276965  | AJ276965 Phycomyce |
| 37 | 21 | 100.0 | 3100 | 8  | PBL278287  | AJ278287 Phycomyce |
| 38 | 21 | 100.0 | 3380 | 11 | G26865     | G26865 human STS S |
| 39 | 21 | 100.0 | 3492 | 9  | S63368     | S63368 Homo sapien |
| 40 | 21 | 100.0 | 3588 | 10 | AB044626   | AB044626 Mus muscu |
| 41 | 21 | 100.0 | 3683 | 6  | AX333705   | AX333705 Sequence  |
| 42 | 21 | 100.0 | 3683 | 6  | AX348016   | AX348016 Sequence  |
| 43 | 21 | 100.0 | 3683 | 6  | AX348018   | AX348018 Sequence  |
| 44 | 21 | 100.0 | 3683 | 6  | AX348020   | AX348020 Sequence  |
| 45 | 21 | 100.0 | 3683 | 9  | HUMNFR     | M32315 Human tumor |

## ALIGNMENTS

RESULT 1  
AX349065  
LOCUS AX349065  
DEFINITION Sequence 3 from Patent WO0202785.  
ACCESSION AX349065  
VERSION AX349065.1 GI:18615100  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct.  
artificial sequence.  
REFERENCE  
1 (sites)  
AUTHORS Sprunck,S., Kluth,A., Becker,D., Luetticke,S. and Loerz,H.  
TITLE Promoters of gene expression in plant carypsops  
JOURNAL Patent: WO 0202785-A 3 10-JAN-2002;  
Aventis CropScience GmbH (DE)  
location/Qualifiers  
1..21  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="Oligonucleotide"

BASE COUNT 16 a 4 c 1 g 0 t  
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 36+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaacaacaacaaca 21  
 ||||||||||||||||||  
 Db 1 AGAACAACAACAACAACA 21

## RESULT 2

RNU76838 116 bp DNA linear ROD 27-SEP-1997  
 LOCUS Rattus norvegicus cyclin A gene, intronic microsatellite repeat  
 DEFINITION region.

ACCESSION U76838 GI:2443415

VERSION U76838.1  
 KEYWORDS Norway rat.  
 SOURCE Rattus norvegicus  
 ORGANISM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE 1 (bases 1 to 116)  
 BROCKMANN, B., SCHNIEDERS, F., KUNZE, B., KRAWCZAK, M. and SCHMIDTKE, J.  
 DNA sequence divergence in a cyclin A gene of Apodemus sylvaticus  
 and Apodemus flavicollis

JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 116)  
 BROCKMANN, B., SCHNIEDERS, F., KUNZE, B., KRAWCZAK, M. and SCHMIDTKE, J.  
 Direct Submission  
 Submitted (01-NOV-1996) Max-Planck-Group, MDC for Molecular  
 Medicine, Robert-Rossle-Str. 10, Berlin-Buch 13125, Germany

FEATURES Location/Qualifiers

source 1..116  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 intron <1..>116  
 /note="cyclin A; Allele: 1"  
 repeat\_region 80..116  
 /note="intronic"  
 /rpt\_family="microsatellite"  
 BASE COUNT 58 a 17 c 19 g 22 t

ORIGIN

Query Match 100.0%; Score 21; DB 10; Length 116;  
 Best Local Similarity 100.0%; Pred. No. 2.36+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaacaacaacaaca 21  
 ||||||||||||||||||  
 Db 94 AGAACAACAACAACAACA 114

## RESULT 3

RNU76839 120 bp DNA linear ROD 27-SEP-1997  
 LOCUS Rattus norvegicus cyclin A gene, intronic microsatellite repeat  
 DEFINITION region.

ACCESSION U76839 GI:2443416

VERSION U76839.1  
 KEYWORDS Norway rat.  
 SOURCE Rattus norvegicus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 120)  
 BROCKMANN, B., SCHNIEDERS, F., KUNZE, B., KRAWCZAK, M. and SCHMIDTKE, J.  
 DNA sequence divergence in a cyclin A gene of Apodemus sylvaticus  
 and Apodemus flavicollis

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 120)  
 BROCKMANN, B., SCHNIEDERS, F., KUNZE, B., KRAWCZAK, M. and SCHMIDTKE, J.  
 Direct Submission  
 Submitted (01-NOV-1996) Max-Planck-Group, MDC for Molecular  
 Medicine, Robert-Rossle-Str. 10, Berlin-Buch 13125, Germany

FEATURES Location/Qualifiers

source 1..120  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 intron <1..>120  
 /note="cyclin A; Allele: 2"  
 repeat\_region 80..120  
 /note="intronic"  
 /rpt\_family="microsatellite"  
 BASE COUNT 61 a 17 c 20 g 22 t

ORIGIN

Query Match 100.0%; Score 21; DB 10; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 2.36+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaacaacaacaaca 21  
 ||||||||||||||||||  
 Db 98 AGAACAACAACAACAACA 118

## RESULT 4

AY023679 224 bp DNA linear. PLN 07-FEB-2001  
 LOCUS Oryza sativa microsatellite MR6004 containing (AAAC)X6, closest to  
 marker R1479, genomic sequence.

ACCESSION AY023679  
 VERSION AY023679.1 GI:12706895

KEYWORDS Oryza sativa.  
 SOURCE Oryza sativa.  
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE 1 (bases 1 to 224)  
 TAO, N., BARBAZUK, W.B., LIU, J., WU, K. and BARRY, G.F.  
 Simple sequence repeats from Monsanto rice genomic sequences  
 Unpublished

JOURNAL 2 (bases 1 to 224)  
 TAO, N., BARBAZUK, W.B., LIU, J., WU, K. and BARRY, G.F.  
 Direct Submission  
 Submitted (10-JAN-2001) Genomics, Monsanto, 800 North Lindbergh  
 Blvd., Creve Coeur, MO 63167, USA

COMMENT Derived from rice genomic sequences generated from the Monsanto  
 Rice Genome Sequencing Project. Please see  
 http://www.rice-research.org for more information. The sequence  
 data were produced primarily in the laboratories of Dr. Leroy Hood  
 at the University of Washington in Seattle.

FEATURES Location/Qualifiers

source 1..224  
 /organism="Oryza sativa"  
 /db\_xref="taxon:4530"  
 repeat\_region 1..224  
 /note="microsatellite MR6004"  
 /rpt\_type=tandem  
 /rpt\_unit=aaac

BASE COUNT 83 a 53 c 45 g 43 t

ORIGIN

Query Match 100.0%; Score 21; DB 8; Length 224;  
 Best Local Similarity 100.0%; Pred. No. 2.46+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaacaacaacaaca 21  
 ||||||||||||||||||  
 Db 99 AGAACAACAACAACAACA 119



RESULT 5  
G35045  
LOCUS  
DEFINITION STS NR03-14s clone NR03-14, sequence tagged site.  
ACCESSION G35045  
VERSION G35045.1 GI:2358168  
KEYWORDS STS; STS sequence; primer; sequence tagged site.  
SOURCE human vector-pbluescript host-XL1 Blue Alu PCR products of chromosome 3 radiation hybrids cloned into pbs.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 239)  
AUTHORS Naylor,S.L., Garcia,D.K., Kerbacher,K., O'Connell,P. and Stanton,V.  
TITLE Human Chromosome 3 STS  
JOURNAL Unpublished  
COMMENT GDB\_DSEG: D3S4038  
Contact: Susan L. Naylor  
San Antonio Genome Center  
Univ. Of TX Health Science Center, San Antonio  
7703 Floyd Curl Dr., San Antonio, TX 78385, USA  
Tel: (210)567-3842  
Fax: (210)567-6781  
Email: naylor@uthscsa.edu  
Primer A: AACACTCATTCTCCTCAGC  
Primer B: CCGCAATTAACCTTCAGTCC  
STS size: 132  
PCR Profile:  
Presoak: 95 degrees C for 5 minutes  
Denaturation: 95 degrees C for 1 minute  
Annealing: 56 degrees C for 1 minute  
Extension: 72 degrees C for 1 minute  
PCR cycles: 30  
Final extension: 72 degrees C for 10 minutes  
Thermal cycler: Perkin Elmer 48  
Protocol:  
Template: 200ng  
Primer: 10ng each  
dNTPs: 4mM of each dNTP/20ul rxn  
Taq polymerase: 1 unit /20ul rxn  
Total Vol: 20ul  
Buffer:  
MgCl2: 1.5mM  
KCl:  
Tris-HCl:  
pH:  
FEATURES  
source  
1..239  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
STS  
83..214  
primer\_bind 83..100  
primer\_bind complement(195..214)  
BASE COUNT 80 a 56 c 39 g 64 t  
ORIGIN  
Query Match 100.0%; Score 21; DB 11; Length 239;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaacaacaacaacaaca 21  
|||||  
Db 11 AGAACAACAACAACAACA 31

RESULT 6  
AU026129/c

LOCUS AU026129 281 bp DNA linear STS 02-MAR-1999  
DEFINITION Rattus norvegicus, OTSUKA clone, OT58.09/882h07, microsatellite  
sequence, sequence tagged site.  
ACCESSION AU026129  
VERSION AU026129.1 GI:4516052  
KEYWORDS STS.  
SOURCE Rattus norvegicus DNA, clone:OT58.09/882h07.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (sites)  
AUTHORS Watanabe,T.K., Hishigaki,H., Kanemoto,N., Miyakita-Mizoguchi,A., Oga,K., Okuno,S., Ono,T., Tsuji,A., Hayashi,H., Adachi,M., Yamasaki,Y., Iriye,Y., Takahashi,E., Takagi,T., Nakamura,Y. and Tanigami,A.  
TITLE The large scale mapping of rat microsatellite markers  
JOURNAL Unpublished (1998)  
REFERENCE 2 (bases 1 to 281)  
AUTHORS Watanabe,T.K.  
TITLE Direct Submission  
JOURNAL Submitted (24-JUL-1998) to the DDBJ/EMBL/GenBank databases. Takeshi K Watanabe, Otsuka Pharmaceutical Co., Ltd., Otsuka GEN Research Institute, 463-10, Kagasuno, Kawachi-cho, Tokushima, Tokushima 771-0192, Japan (E-mail:watanabe@otsuka.genome.ad.jp, Tel:81-886-65-2888, Fax:81-886-37-1035)  
FEATURES  
source  
1..281  
Location/Qualifiers  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="OT58.09/882h07"  
/note="OT58.09/882h07F=5'-CGATCCATCTCCGCGTCTC-3', OT58.09/882h07R=5'-GTCACTCTCACTCCACAGTG-3'"

BASE COUNT 42 a 42 c 71 g 126 t  
ORIGIN

Query Match 100.0%; Score 21; DB 11; Length 281;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaacaacaacaacaaca 21  
|||||  
Db 128 AGAACAACAACAACAACA 108

RESULT 7  
HUMRPTPLM 309 bp DNA linear PRI 11-MAR-1996  
LOCUS  
DEFINITION Human microsatellite dinucleotide repeat polymorphism.  
ACCESSION L20486  
VERSION L20486.1 GI:1220359  
KEYWORDS microsatellite polymorphism; microsatellite repeat; repeat polymorphism.  
SOURCE Homo sapiens (tissue library: Bluescript) DNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Sainz,J. and Pulst,S.-M.  
JOURNAL Unpublished (1993)  
FEATURES  
source  
1..309  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="22q12"  
/tissue\_lib="Bluescript"  
212..248  
/note="polymorphism"  
/rpt\_family="microsatellite"  
/rpt\_type=tandem

repeat\_region

BASE COUNT 113 a 82 c 52 g 62 t  
ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 309;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agaaacaacaacaacaaca 21  
 ||||||||||||||||||  
 Db 41 AGAAACAACAACAACAACA 61

RESULT 8  
 HMDUT925 381 bp DNA linear STS 26-JUL-1993  
 LOCUS Human chromosome 10 STS UT925, sequence tagged site.  
 DEFINITION L16357  
 ACCESSION L16357.1 GI:308719  
 VERSION STS; PCR primer; microsatellite marker; microsatellite repeat;  
 KEYWORDS repeat polymorphism; sequence tagged site; tetranucleotide repeat.  
 SOURCE Homo sapiens DNA.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 381)  
 Gerken,S.C., Matsunami,N., Lawrence,E., Carlson,M., Moore,M.,  
 Ballard,L., Melis,R., Robertson,M., Bradley,P., Elsnier,T.,  
 Tingey,A., Rodriguez,P., Albertsen,H., Lalouel,J.-M. and White,R.  
 Genetic and physical mapping of simple sequence repeat containing  
 sequence tagged sites from the human genome  
 Unpublished (1993)

JOURNAL  
 COMMENT Submitted by: Utah Center for Human Genome Research University of  
 Utah, Dept. of Human Genetics  
 2160 Eccles Institute of Human Genetics  
 Salt Lake City, UT 84112  
 e-mail: sts@corona.med.utah.edu  
 Primer A: GCACCTCCACCCATGCAAC  
 Primer B: GTCCCTGTGAGAACTGGATG  
 32P-label: A Primer  
 PCR Profile:  
 Initial Denaturation: 94C 300sec  
 PCR Cycles: 5  
 Denaturation: 94C 10sec  
 Annealing: 60C 10sec  
 Extension: 72C 20sec  
 Mg++: 1mM  
 Gel: Acrylamide 7%, Formamide 32%, Urea 34%  
 Alleles: 7.

FEATURES  
 source Location/Qualifiers  
 1..381  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="10"  
 STS 49..223  
 /standard\_name="STS UT925"  
 primer\_bind 49..67  
 primer\_bind complement(203..223)  
 BASE COUNT 156 a 82 c 94 g 48 t 1 others  
 ORIGIN

Query Match 100.0%; Score 21; DB 11; Length 381;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agaaacaacaacaacaaca 21  
 ||||||||||||||||||  
 Db 171 AGAAACAACAACAACAACA 191

RESULT 9  
 HMDUT1400 407 bp DNA linear STS 26-JUL-1993  
 LOCUS Human STS UT1400, sequence tagged site.  
 DEFINITION

ACCESSION L17810  
 VERSION L17810.1 GI:307674  
 KEYWORDS STS; PCR primer; STS sequence; microsatellite marker;  
 microsatellite repeat; repeat polymorphism; sequence tagged site.  
 SOURCE Homo sapiens DNA.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 407)  
 Gerken,S.C., Matsunami,N., Lawrence,E., Carlson,M., Moore,M.,  
 Ballard,L., Melis,R., Robertson,M., Bradley,P., Elsnier,T.,  
 Tingey,A., Rodriguez,P., Albertsen,H., Lalouel,J.-M. and White,R.  
 Genetic and physical mapping of simple sequence repeat containing  
 sequence tagged sites from the human genome  
 Unpublished (1993)

JOURNAL  
 COMMENT Submitted by: Utah Center for Human Genome Research University of  
 Utah, Dept. of Human Genetics  
 2160 Eccles Institute of Human Genetics  
 Salt Lake City, UT 84112  
 e-mail: sts@corona.med.utah.edu  
 Primer A: GATGGAGAGACTGCTTAAG  
 Primer B: TGGCATGACCTCCTGAGTAG  
 32P-label: B Primer  
 PCR Profile:  
 Initial Denaturation: 94C 300sec  
 PCR Cycles: 5  
 Denaturation: 94C 10sec  
 Annealing: 62C 10sec  
 Extension: 72C 20sec  
 Mg++: 1mM  
 Gel: Acrylamide 7%, Formamide 32%, Urea 34%  
 Alleles: 1.

FEATURES  
 source Location/Qualifiers  
 1..407  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 STS 68..362  
 /standard\_name="STS UT1400"  
 primer\_bind 68..87  
 primer\_bind complement(343..362)  
 BASE COUNT 166 a 63 c 109 g 61 t 8 others  
 ORIGIN

Query Match 100.0%; Score 21; DB 11; Length 407;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agaaacaacaacaacaaca 21  
 ||||||||||||||||||  
 Db 273 AGAAACAACAACAACAACA 293

RESULT 10  
 HMD8S320T 497 bp DNA linear PRI 08-OCT-1993  
 LOCUS Homo sapiens (D8S320 locus) DNA sequence, tetranucleotide repeat  
 DEFINITION polymorphism.  
 ACCESSION L12265.1 GI:388013  
 VERSION L12265  
 KEYWORDS tetranucleotide repeat polymorphism.  
 SOURCE Homo sapiens DNA.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 497)  
 Riley,R., Nelson,L., Lu,J., Robertson,M., Ballard,L., Connolly,J.  
 and Ward,K.  
 Tetranucleotide repeat polymorphism at the D8S320 locus  
 Hum. Mol. Genet. 2, 1512-1512 (1993)

JOURNAL  
 COMMENT Location/Qualifiers  
 1..497  
 /organism="Homo sapiens"

```

/isolate="KW205"
/db_xref="taxon:9606"
/clone_lib="flow-sorted chromosome 8 specific cosmid
library (LA08MC01)"
1..497
/standard_name="D8S320 locus"
repeat_region

BASE COUNT      270 a      45 c      105 g      62 t      15 others
ORIGIN

Query Match      100.0%; Score 21; DB 9; Length 497;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1      agaaacaacaacaacaaca 21
        |||||||
Db      287      AGAACAACAACAACAACAACA 307

RESULT 11
LOCUS      HUMUT5307      497 bp      DNA      linear      STS 26-JUL-1993
DEFINITION      Human chromosome 8 STS UT5307, sequence tagged site.
ACCESSION      U18590.1 GI:308301
VERSION      U18590.1 GI:308301
KEYWORDS      STS; PCR primer; STS sequence; microsatellite marker;
SOURCE      microsatellite repeat; repeat polymorphism; sequence tagged site.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 497)
Gerken,S.C., Matsunami,N., Lawrence,E., Carlson,M., Moore,M.,
Ballard,L., Mellis,R., Robertson,M., Bradley,P., Elmer,T.,
Tingey,A., Rodriguez,P., Albertsen,H., Lalouel,J.-M. and White,R.
Genetic and physical mapping of simple sequence repeat containing
sequence tagged sites from the human genome
Unpublished (1993)
Submitted by: Utah Center for Human Genome Research University of
Utah, Dept. of Human Genetics
2160 Eccles Institute of Human Genetics
Salt Lake City, UT 84112
e-mail: sts@corona.med.utah.edu
Primer A: GGATTTACTGCAATGCAAAAC
Primer B: GTCAATTCAGAGACTGTAAACAG
PCR Profile:
Initial Denaturation: 94C 300sec
PCR Cycles: 5
Denaturation: 94C 10sec
Annealing: 60C 10sec
Extension: 72C 20sec
Mg++: 2mM
Gel: Acrylamide 7%, Formamide 32%, Urea 34%
Alleles: 4.

FEATURES
source      1..497
            location/Qualifiers
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /map="8"
STS      38..449
            /standard_name="STS UT5307"
            primer_bind      38..62
            primer_bind      complement(425..449)
BASE COUNT      270 a      45 c      105 g      62 t      15 others
ORIGIN

Query Match      100.0%; Score 21; DB 11; Length 497;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1      agaaacaacaacaacaaca 21

```

```

Db      287      AGAACAACAACAACAACAACA 307
        |||||||
RESULT 12
LOCUS      G24380/c      503 bp      DNA      linear      STS 31-MAY-1996
DEFINITION      human STS WI-31303, sequence tagged site.
ACCESSION      G24380
VERSION      G24380.1 GI:1344706
KEYWORDS      STS; STS sequence; primer; sequence tagged site.
SOURCE      human STS derived from sequences in dbEST and the UniGene
collection.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 503)
Hudson,T.
Whitehead Institute/MIT Center for Genome Research; Physically
Mapped STS
Unpublished

JOURNAL
COMMENT
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu

Primer A: AAGCCCTGTAACCTGAGCT
Primer B: GGCAGAACATCTCGAATA
STS size: 150
PCR Profile:
Presoak:
Denaturation:
Annealing: 56 degrees C
Polymerization:
PCR Cycles: 35
Thermal Cycler:
Protocol:
Template: 10 ng
Primer: each 5 pM
dNTPs: each 4 mM
Tag polymerase: 0.025 units/ul
Total Vol: 20 ul
Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCL: 10 mM
pH: 9.3

Derived from dbEST (genbank accession H92026).
location/Qualifiers
1..503
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="183.2 CR from top of Chr7 linkage group"
STS      1..150
            primer_bind      1..21
            primer_bind      complement(131..150)
BASE COUNT      145 a      97 c      118 g      128 t      15 others
ORIGIN

Query Match      100.0%; Score 21; DB 11; Length 503;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1      agaaacaacaacaacaaca 21
        |||||||
Db      181      AGAACAACAACAACAACAACA 161

```

RESULT 13  
LOCUS G42246/c 552 bp DNA linear STS 29-NOV-1999  
DEFINITION 58p18c1 Human PAC clone Homo sapiens STS genomic, sequence tagged site.  
ACCESSION G42246  
VERSION G42246.1 GI:4115876  
KEYWORDS STS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 552)  
AUTHORS Paavola,P., Avela,K., Horelli-Kuitunen,N., Barlund,M., Kallioniem,A., Idenheimo,N., Kyttälä,M., de la Chapelle,A., Palotie,A., Lehesjoki,A.-E. and Peltonen,L.  
High-resolution physical and genetic mapping of the critical region for Meckel syndrome and Mulibrey Nanism on chromosome 17q22-q23  
Genome Res. 9 (3), 267-276 (1999)  
JOURNAL 99177349  
MEDLINE 10077533  
COMMENT Contact: Kristina Avela  
Department of Medical Genetics  
University of Helsinki  
Helsinki, Finland  
Email: Kristina.Avela@helsinki.fi  
Primer A: TATGTCAGATGTCAGTCG  
Primer B: CCTCATTTATCTGACCAAGCC  
STS size: 107.  
Location/Qualifiers  
1..552  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Human PAC clone"  
/lab\_host="E. coli DH10B"  
/note="vector: pCYPAC-1; V-type: pl-derived vector"

STS  
primer\_bind 111..132  
BASE COUNT 152 a 107 c 127 g 165 t 1 others  
ORIGIN

Query Match 100.0%; Score 21; DB 11; Length 552;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaacaacaacaacaa 21  
Db 338 AGAACAACAACAACAACA 318

RESULT 14  
LOCUS HUMUT6478 561 bp DNA linear STS 29-DEC-1994  
DEFINITION Human STS UT6478, sequence tagged site.  
ACCESSION L30582  
VERSION L30582.1 GI:604976  
KEYWORDS STS; PCR primer; STS sequence; microsatellite DNA; microsatellite marker; sequence tagged site; tetranucleotide repeat.  
SOURCE Homo sapiens DNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 561)  
AUTHORS Gerken,S.C., Matsunami,N., Plaelke,R., Albertsen,H., Ballard,L., Melis,R., Lawrence,E., Moore,M., Holik,P.R., Carlson,M., Zhao,X., Robertson,M., Bradley,P., Elsnier,T., Tingey,A., Lalouel,J.-M. and White,R.  
Genetic and physical mapping of simple sequence repeat containing

JOURNAL Unpublished (1994)  
COMMENT Submitted by: Utah Center for Human Genome Research University of Utah, Dept. of Human Genetics  
2160 Eccles Institute of Human Genetics  
Salt Lake City, UT 84112  
e-mail: sts@corona.med.utah.edu  
Primer A: TACATCCAGCTGGGCAC  
Primer B: CTCCTCAGAGTCTCCATCG  
End to Label: Primer B

PCR profile:  
Initial Denaturation: 94C 30sec  
Cycles Denaturation Annealing Extension 5 94  
C 10 sec. 64 C 10 sec. 72 C 20 sec. 30  
60 C 10 sec. 72 C 20 sec. Mg++: 1.00 mM  
Gel: Acrylamide 7%, Formamide 32%, Urea 34%  
Alleles: 9.

FEATURES  
source Location/Qualifiers  
1..561  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="2"

STS  
primer\_bind 167..343  
primer\_bind 167..186  
primer\_bind complement(324..343)  
BASE COUNT 170 a 153 c 135 g 93 t 10 others  
ORIGIN

Query Match 100.0%; Score 21; DB 11; Length 561;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaacaacaacaacaa 21  
Db 269 AGAACAACAACAACAACA 289

RESULT 15  
LOCUS AF254660/c 617 bp DNA linear VRT 02-MAY-2001  
DEFINITION Tetrao urogallus microsatellite TWT8 sequence.  
ACCESSION AF254660  
VERSION AF254660.1 GI:13926063  
KEYWORDS western capercaillie.  
SOURCE Tetrao urogallus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauaria; Aves; Neognathae; Galliformes; Tetraonidae; Tetrao.

REFERENCE 1 (bases 1 to 617)  
AUTHORS Segelbacher,G., Paxton,R. and Steinbreck,G.  
TITILE Characterization of hypervariable microsatellites in Capercaillie (Tetrao urogallus)  
JOURNAL Unpublished  
DEFINITION 2 (bases 1 to 617)  
AUTHORS Segelbacher,G.  
TITILE Direct Submision  
JOURNAL Submitted (12-APR-2000) Unit of Wildlife Research and Management, TU Munich, Am Hochanger 13, Freising 85354, Germany

FEATURES  
source Location/Qualifiers  
1..617  
/organism="Tetrao urogallus"  
/db\_xref="taxon:100830"  
repeat\_region 1..617  
/note="microsatellite TWT8"  
/rpt\_type=tandem

BASE COUNT 62 a 156 c 69 g 330 t  
ORIGIN

Thu Aug 1 08:29:59 2002

us-09-899-718a-3.rge

Page 7

Query Match 100.0%; Score 21; DB 5; Length 617;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 agaaacaaacaaacaaacaa 21  
|||||  
Db 58 AGAAACAAACAAACAAACAA 38

Search completed: July 31, 2002, 14:00:51  
Job time: 17493 sec





Db 2363 CGCAATTACGTCGCGCCAGCAGC 2387

RESULT 2

US-08-162-809-13

Sequence 13, Application US/08162809

Patent No. 5457048

GENERAL INFORMATION:

APPLICANT: Pasquale, Elena B.

APPLICANT: Sajjad, Fereydoon G.

TITLE OF INVENTION: NOVEL EPH-RELATED TYROSINE KINASES,

TITLE OF INVENTION: NUCLEOTIDE SEQUENCES, AND METHODS OF USE

NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:

ADDRESSEE: CAMPBELL AND FLORES

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

STATE: California

COUNTRY: United States of America

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/162,809

FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-LJ 9503

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001

TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 3591 base pairs

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: linear

FEATURE:

NAME/KEY: CDS

LOCATION: 2..2965

US-08-162-809-13

Query Match 64.1%; Score 18.6; DB 1; Length 3591;

Best Local Similarity 84.0%; Pred. No. 18;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 cgcgaagcgcgctgcgcagccagc 27

Db 2408 CGCAATTACGTCGCGCCAGCAGC 2432

US-09-095-443-1

Sequence 1, Application US/09095443

Patent No. 6342593

GENERAL INFORMATION:

APPLICANT: Plowman, Gregory

APPLICANT: Peles, Eliot

TITLE OF INVENTION: DIAGNOSIS AND TREATMENT

TITLE OF INVENTION: OF ALP RELATED DISORDERS

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Lyon & Lyon

US-09-103-840A-2

Sequence 2, Application US/09103840A

Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

TITLE OF INVENTION: TUBERCULOSIS

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 2

LENGTH: 4403765

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

FEATURE:

US-09-103-840A-2

Sequence 2, Application US/09103840A

Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

TITLE OF INVENTION: TUBERCULOSIS

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 2

LENGTH: 4403765

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

FEATURE:

US-09-095-443-1

Query Match 62.1%; Score 18; DB 4; Length 4456;

Best Local Similarity 80.8%; Pred. No. 31;

Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 gcaaaagcgcgctgcgcagccagc 29

Db 1846 GCCATGCGGTCGCGCCAGCAGC 1871

US-09-103-840A-2

Sequence 2, Application US/09103840A

Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

TITLE OF INVENTION: TUBERCULOSIS

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 2

LENGTH: 4403765

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

FEATURE:

OTHER INFORMATION: CDC 1551

OTHER INFORMATION: "n" bases at various positions throughout the sequence

US-09-103-840A-2

Query Match 60.0%; Score 17.4; DB 4; Length 4403765;

Best Local Similarity 75.9%; Pred. No. 38;

Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;







|                        |                 |                       |           |               |
|------------------------|-----------------|-----------------------|-----------|---------------|
| Query Match:           | 57.9%;          | Score 16.8;           | DB 3;     | Length 36506; |
| Best Local Similarity: | 90.0%;          | Pred. No. 1e+02;      |           |               |
| Matches 18;            | Conservative 0; | Mismatches 2;         | Indels 0; | Gaps 0;       |
| QY                     | 9               | ggcgcgtgcgcgcgcacaca  | 28        |               |
|                        |                 |                       |           |               |
| db                     | 10958           | ggcgcgttcgcgcgcgcgcga | 10939     |               |

RESULT 14  
US-08-913-362-3  
; Sequence 3, Application US/08913362  
; Patent No. 6287574  
GENERAL INFORMATION:  
APPLICANT: Brodeur, Bernard R  
APPLICANT: Martin, Denis  
APPLICANT: Hamel, Josee  
APPLICANT: Rioux, Clement  
TITLE OF INVENTION: PROTEINASE K RESISTANT SURFACE PROTEIN  
TITLE OF INVENTION: OF NEISSERIA MENINGITIDIS  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,362  
FILING DATE: 13-NOV-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/406,362  
FILING DATE: 17-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/001,983  
FILING DATE: 04-AUG-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 047998/0128  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 710 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Neisseria meningitidis  
STRAIN: MCH88  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 116..643  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 116..172  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 173..643  
US-08-913-362-3

Query Match 56.6%; Score 16.4; DB 4; Length 710;  
Best Local Similarity 76.9%; Pred. No. 1.4e+02;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
OY 4 gcaaaagcgctgcgcagccacgac 29  
||| ||||||||| |||||

Db 363 GCATCGCGCGGTCGCCATTACGAC 388  
RESULT 15  
US-08-913-362-1  
; Sequence 1, Application US/08913362  
; Patent No. 6287574  
GENERAL INFORMATION:  
APPLICANT: Brodeur, Bernard R  
APPLICANT: Martin, Denis  
APPLICANT: Hamel, Josee  
APPLICANT: Rioux, Clement  
TITLE OF INVENTION: PROTEINASE K RESISTANT SURFACE PROTEIN  
TITLE OF INVENTION: OF NEISSERIA MENINGITIDIS  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,362  
FILING DATE: 13-NOV-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/406,362  
FILING DATE: 17-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/001,983  
FILING DATE: 04-AUG-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 047998/0128  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 830 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Neisseria meningitidis  
STRAIN: 608B  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 143..667  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 143..199  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 200..667  
US-08-913-362-1

Query Match 56.6%; Score 16.4; DB 4; Length 830;  
Best Local Similarity 76.9%; Pred. No. 1.4e+02;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
OY 4 gcaaaagcgctgcgcagccacgac 29

Thu Aug 1 08:29:58 2002

us-09-899-718a-2.rni

Page 7

Db 387 GCATCGGCGCGCTCCGCCATTACGAC 412

Search completed: July 31, 2002, 12:20:34  
Job time: 11657 sec

1  
:

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:11:51 ; Search time 5855.71 Seconds  
(without alignments)  
66.843 Million cell updates/sec

Title: US-09-899-718a-2

Perfect score: 29  
Sequence: 1 caccgcaagcgccgctgcgcacgacacac 29

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estio:\*  
8: em\_estc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_iny:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| C 1        | 19.8  | 68.3        | 222    | 10    | N46230 YY72B11.r1  |
| C 2        | 19.8  | 68.3        | 255    | 9     | AA227176 ZR20G09.r |
| C 3        | 19.8  | 68.3        | 350    | 10    | M94459 ZEL2E03.r1  |
| C 4        | 19.8  | 68.3        | 984    | 10    | BG024310 602274479 |
| C 5        | 19.8  | 68.3        | 2306   | 11    | BC004149 Homo sapi |
| C 6        | 19.4  | 66.9        | 1028   | 10    | BE733309 601569494 |
| C 7        | 19.4  | 66.9        | 1100   | 10    | BE893877 601436242 |
| C 8        | 19.4  | 66.9        | 1829   | 12    | AO782504 HS.3174.A |
| C 9        | 19.2  | 66.2        | 325    | 10    | DA8137 R1CS14202A  |
| C 10       | 19.2  | 66.2        | 1096   | 12    | CNS04XLD           |
| C 11       | 19.2  | 65.5        | 290    | 9     | AM972385 EST384476 |
| C 12       | 19.2  | 65.5        | 326    | 9     | AV433974 AV433974  |
| C 13       | 19.2  | 65.5        | 416    | 9     | AV435705 AV435705  |
| C 14       | 19.2  | 65.5        | 452    | 9     | AV430263 AV430263  |
| C 15       | 19.2  | 65.5        | 479    | 9     | BB855054 BB855054  |
| C 16       | 19.2  | 65.5        | 509    | 9     | AV435479 AV435479  |
| C 17       | 19.2  | 65.5        | 517    | 9     | AV433024 AV433024  |

|      |      |      |      |    |          |          |
|------|------|------|------|----|----------|----------|
| C 18 | 19   | 65.5 | 519  | 9  | AV435216 | AV435216 |
| C 19 | 19   | 65.5 | 519  | 9  | AV435485 | AV435485 |
| C 20 | 19   | 65.5 | 528  | 9  | AV432179 | AV432179 |
| C 21 | 19   | 65.5 | 528  | 9  | AV434482 | AV434482 |
| C 22 | 19   | 65.5 | 529  | 9  | AV432635 | AV432635 |
| C 23 | 19   | 65.5 | 530  | 9  | AV433071 | AV433071 |
| C 24 | 19   | 65.5 | 530  | 9  | AV435732 | AV435732 |
| C 25 | 19   | 65.5 | 532  | 9  | AV432358 | AV432358 |
| C 26 | 19   | 65.5 | 533  | 9  | AV434711 | AV434711 |
| C 27 | 19   | 65.5 | 534  | 9  | AV432450 | AV432450 |
| C 28 | 19   | 65.5 | 534  | 9  | AV433137 | AV433137 |
| C 29 | 19   | 65.5 | 534  | 9  | AV434241 | AV434241 |
| C 30 | 19   | 65.5 | 536  | 9  | AV435646 | AV435646 |
| C 31 | 19   | 65.5 | 544  | 9  | AV434121 | AV434121 |
| C 32 | 19   | 65.5 | 546  | 9  | AV432203 | AV432203 |
| C 33 | 19   | 65.5 | 548  | 9  | AV434539 | AV434539 |
| C 34 | 19   | 65.5 | 549  | 9  | AV432005 | AV432005 |
| C 35 | 19   | 65.5 | 549  | 9  | AV437069 | AV437069 |
| C 36 | 19   | 65.5 | 552  | 9  | AV431821 | AV431821 |
| C 37 | 19   | 65.5 | 552  | 10 | BG274284 | BG274284 |
| C 38 | 19   | 65.5 | 571  | 10 | BF258049 | BF258049 |
| C 39 | 19   | 65.5 | 789  | 10 | BG716721 | BG716721 |
| C 40 | 19   | 65.5 | 917  | 10 | BF213345 | BF213345 |
| C 41 | 18.6 | 64.1 | 1166 | 11 | AK021174 | AK021174 |
| C 42 | 18.6 | 64.1 | 301  | 12 | A2922255 | A2922255 |
| C 43 | 18.6 | 64.1 | 500  | 10 | BE587380 | BE587380 |
| C 44 | 18.6 | 64.1 | 684  | 12 | AQ329925 | AQ329925 |
| C 45 | 18.6 | 64.1 | 793  | 10 | BG923682 | BG923682 |

#### ALIGNMENTS

RESULT 1  
LOCUS N46230/c 222 bp mRNA linear EST 14-FEB-1996  
DEFINITION YY72B11.r1 Soares\_multiple\_sclerosis\_2NBHMSP Homo sapiens cDNA  
clone IMAGE:279069 5', mRNA sequence.

ACCESSION N46230  
VERSION N46230.1 GI:1187396

KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 222)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman

AUTHORS 'M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,  
Rifkin, L., Rohlfing, T., Soares, M., Tan, P., Trevaskis, E., Waterston

, R., Williamson, A., Wohlmann, P. and Wilson, R.  
The WashU-Merck EST Project

Unpublished (1995)

CONTACT: Wilson RK  
Washington University School of Medicine

444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800

Fax: 314 286 1810

Email: estevatson.wustl.edu

This clone is available royalty-free through LIND ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: T7

High quality sequence stop: 185.

#### FEATURES

source  
1..222  
/organism="Homo sapiens"  
/db\_xref="GDB:3897445"  
/db\_xref="taxon:9606"  
/clone="IMAGE:279069"  
/clone\_lib="Soares\_multiple\_sclerosis\_2NBHMSP"  
/sex="male"  
/tissue\_type="multiple sclerosis lesions"  
/dev\_stage="Age 46"  
/lab\_host="DH10B (ampicillin resistant)"

/note="Vector: pT73D (Pharmacia) with a modified polylinker V-type: phagemid; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5' TGTACCATCTGGAAGTGGAGCGCCGCGATTTTTTTTTTTTTTTT 3'), double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA from 4 multiple sclerosis lesions from one patient was kindly provided by Dr. Kevin G. Becker (NINDS/NIH).

BASE COUNT 43 a 71 c 64 g 38 t 6 others

ORIGIN

Query Match 68.3%; Score 19.8; DB 10; Length 222;  
Best Local Similarity 91.3%; Pred. No. 4.5e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 gcaaaagcgcgtcgccagccac 26  
|||||  
Db 151 GCAAAAGCGCGTCGCGCGGCAC 129

RESULT 2 AA227176 255 bp mRNA linear EST 11-MAR-1998  
LOCUS AA227176 zr20g09.r1 stratagene NT2 neuronal precursor 937230 Homo sapiens  
DEFINITION cDNA clone IMAGE:664000 5', mRNA sequence.  
AA227176  
ACCESSION AA227176.1 GI:1848846  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 255)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisler, G., Jost, S., Kitzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steppe, M., Tan, F., Theisling, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
White, Y., Wylie, T., Waterston, R. and Wilson, R.  
WashU-NCI human EST Project  
Unpublished (1997)  
TITLE JOURNAL  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1138 Std Error: 0.00  
Seq primer: -28ml3 rev1 ET from Amerisham  
High quality sequence stop: 247.  
Location/Qualifiers

FEATURES  
SOURCE 1..255  
/organism="Homo sapiens"  
/db\_xref="GDB:5425859"  
/db\_xref="taxon:9606"  
/clone="IMAGE:664000"  
/clone\_lib="Stratagene NT2 neuronal precursor 937230"  
/tissue\_type="neuroepithelial cells"  
/dev\_stage="Ntera-2 neuroepithelial cells"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: brain; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: oligo dT. Uninduced, exponentially growing neuroepithelial cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;  
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTTCGCGCAGCAG 3' -3' adaptor sequence: 5' CTCGAGTTTGTTTTGTTTT 3' "

BASE COUNT 44 a 71 c 75 g 65 t

## ORIGIN

Query Match 68.3%; Score 19.8; DB 9; Length 255;  
Best Local Similarity 91.3%; Pred. No. 4.7e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 gcaaaagcgcgtcgccagccac 26  
|||||  
Db 97 GCAAAAGCGCGTCGCGCGGCAC 75

RESULT 3 W94459 350 bp mRNA linear EST 16-JUL-1996  
LOCUS W94459 zsl2e03.r1 Soares fetal heart.NbHH19W Homo sapiens cDNA clone  
DEFINITION IMAGE:358780 5', mRNA sequence.  
W94459  
ACCESSION W94459.1 GI:1423589  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 350)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, N., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)  
TITLE JOURNAL  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: mb. REGA+ET  
High quality sequence stop: 321.  
Location/Qualifiers

FEATURES  
SOURCE 1..350  
/organism="Homo sapiens"  
/db\_xref="GDB:1275324"  
/db\_xref="taxon:9606"  
/clone="IMAGE:358780"  
/clone\_lib="Soares\_fetal\_heart\_NbHH19W"  
/sex="unknown"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5' TGTACCATCTGGAAGTGGAGCGCCGCGATTTTTTTTTTTTTTTT 3'), double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M. Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NbHH19W."

BASE COUNT 65 a 107 c 100 g 76 t 2 others

Query Match 68.3%; Score 19.8; DB 10; Length 350;  
Best Local Similarity 91.3%; Pred. No. 5.1e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 gcaaaagcgcgtcgccagccac 26  
|||||



Db 195 GCAAGGCGCGTCGGCGGCAC 173

RESULT 4  
LOCUS BG024310/c  
DEFINITION 602274479F1 NIH\_MGC\_85 Homo sapiens cDNA clone IMAGE:4362450 5', mRNA sequence.  
ACCESSION BG024310  
VERSION BG024310.1 GI:12409756  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: Louis Staudt, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Plate: LHAM10006 row: n column: 19

FEATURES  
source  
Location/Qualifiers  
1..984  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4362450"  
/clone\_1ib="NIH\_MGC\_85"  
/tissue\_type="lymphoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lymph; Vector: PCMV-SPORE6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC library."

BASE COUNT 192 a 356 c 249 g 187 t

ORIGIN

Query Match 68.3%; Score 19.8; DB 10; Length 984;  
Best Local Similarity 91.3%; Pred. No. 6.9e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 gcaaaagcgctcgccagccac 26  
|||||  
MGC|||||

Db 152 GCAAGGCGCGTCGGCGGCAC 130

RESULT 5  
LOCUS BC004149/c  
DEFINITION BC004149 Homo sapiens, clone IMAGE:2819393, mRNA.  
ACCESSION BC004149  
VERSION BC004149.1 GI:14708750  
KEYWORDS HTC.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Strausberg, R.  
TITLE Direct Submision  
JOURNAL Submitted (01-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

REMARK  
COMMENT  
USA  
NIH-MGC Project URL: http://mgc.nci.nih.gov  
Contact: MGC help desk  
Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: DCTD/DRP  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Institute for Systems Biology  
http://www.systemsbio.org  
Contact: amadan@systemsbiology.org  
Anup Madan, Rachel Dickhoff, Jessica Fahay, Stephanie Ford, Julia Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Series: IRAL Plate: 2 Row: 0 Column: 3  
This clone has the following problem: frame shifted.

FEATURES  
source  
Location/Qualifiers  
1..2306  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2819393"  
/tissue\_type="lung, small cell carcinoma"  
/clone\_1ib="NIH\_MGC\_7"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"

BASE COUNT 460 a 764 c 640 g 442 t

ORIGIN

Query Match 68.3%; Score 19.8; DB 11; Length 2306;  
Best Local Similarity 91.3%; Pred. No. 8.8e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 gcaaaagcgctcgccagccac 26  
|||||  
MGC|||||

Db 1354 GCAAGGCGCGTCGGCGGCAC 1332

RESULT 6  
LOCUS BE733309  
DEFINITION 601569494F1 NIH\_MGC\_21 Homo sapiens cDNA clone IMAGE:3844366 5', mRNA sequence.  
ACCESSION BE733309  
VERSION BE733309.1 GI:10147301  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LCM540 row: k column: 23

High quality sequence stop: 136.

FEATURES  
source  
Location/Qualifiers  
1..1028  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3844366"  
/clone\_1ib="NIH\_MGC\_21"  
/tissue\_type="choriocarcinoma"  
/lab\_host="DH10B (phage-resistant)"





JOURNAL DNA Res. 7, 223-227 (2000)  
 MEDLINE 20363100  
 COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
 source  
 1. 326  
 /organism="Porphyra yezoensis"  
 /strain="TU-1"  
 /db\_xref="taxon:2788"  
 /clone="PM036f02.r"  
 /clone\_lib="Porphyra yezoensis TU-1"  
 /note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 25 a 147 c 78 g 76 t  
 ORIGIN

Query Match 65.5%; Score 19; DB 9; Length 326;  
 Best Local Similarity 81.5%; Pred. No. 1e+03; 5; Indels 0; Gaps 0;  
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 cgcgaagcgctgcgcgcgcgcgcac 29  
 |||||  
 Db 264 CGCACAGCGCGACGCGACCGACGAC 238

RESULT 13  
 AVA35705/c 416 bp mRNA linear EST 23-AUG-2000  
 LOCUS AVA35705 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone  
 DEFINITION PM061c04\_r 5', mRNA sequence.  
 ACCESSION AVA35705  
 VERSION AVA35705.1 GI:8590930  
 KEYWORDS EST.  
 SOURCE Porphyra yezoensis.  
 ORGANISM Porphyra yezoensis.  
 Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;  
 Porphyra.

REFERENCE  
 AUTHORS 1 (bases 1 to 416)  
 Nikaido, I., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N. and  
 Tabata, S.  
 TITLE Generation of 10,154 expressed sequence tags from a leafy  
 gametophyte of a marine red alga, Porphyra yezoensis  
 JOURNAL DNA Res. 7, 223-227 (2000)  
 MEDLINE 20363100  
 COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
 source  
 1. 416  
 /organism="Porphyra yezoensis"  
 /strain="TU-1"  
 /db\_xref="taxon:2788"  
 /clone="PM061c04\_r"  
 /clone\_lib="Porphyra yezoensis TU-1"  
 /note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 47 a 178 c 104 g 87 t  
 ORIGIN

Query Match 65.5%; Score 19; DB 9; Length 416;  
 Best Local Similarity 81.5%; Pred. No. 1e+03;  
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 cgcgaagcgctgcgcgcgcgcgcac 29  
 |||||  
 Db 276 CGCACAGCGCGACGCGACCGACGAC 250

RESULT 14  
 AVA30263/c 452 bp mRNA linear EST 23-AUG-2000  
 LOCUS AVA30263 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone  
 DEFINITION PL015h12\_r 5', mRNA sequence.  
 ACCESSION AVA30263  
 VERSION AVA30263.1 GI:8585488  
 KEYWORDS EST.  
 SOURCE Porphyra yezoensis.  
 ORGANISM Porphyra yezoensis.  
 Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;  
 Porphyra.

REFERENCE  
 AUTHORS 1 (bases 1 to 452)  
 Nikaido, I., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N. and  
 Tabata, S.  
 TITLE Generation of 10,154 expressed sequence tags from a leafy  
 gametophyte of a marine red alga, Porphyra yezoensis  
 JOURNAL DNA Res. 7, 223-227 (2000)  
 MEDLINE 20363100  
 COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
 source  
 1. 452  
 /organism="Porphyra yezoensis"  
 /strain="TU-1"  
 /db\_xref="taxon:2788"  
 /clone="PL015h12\_r"  
 /clone\_lib="Porphyra yezoensis TU-1"  
 /note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 48 a 186 c 119 g 99 t  
 ORIGIN

Query Match 65.5%; Score 19; DB 9; Length 452;  
 Best Local Similarity 81.5%; Pred. No. 1e+03;  
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 cgcgaagcgctgcgcgcgcgcgcac 29  
 |||||  
 Db 267 CGCACAGCGCGACGCGACCGACGAC 241

RESULT 15  
 BB855054/c 479 bp mRNA linear EST 26-NOV-2001  
 LOCUS BB855054 RIKEN full-length enriched, B16 F10Y cells Mus musculus  
 DEFINITION cDNA clone G370022K07 5', mRNA sequence.  
 ACCESSION BB855054  
 VERSION BB855054.1 GI:17096508  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Mammalia; Eutheria; Chordata; Vertebrata; Euteleostomi;  
 Eukaryota; Metazoa; Chordata; Vertebrata; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS 1 (bases 1 to 479)  
 Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,  
 Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Imetani, K., Ishii,  
 Y., Ito, M., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Matsuyama, T.,  
 Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T.,  
 Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,  
 Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa,  
 A., Takehashi, F., Takaku-Akahita, S., Tanaka, T., Tomaru, A., Toya, T.,  
 Wachihi, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.,  
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.  
 2001)  
 JOURNAL unpublished (2001)  
 COMMENT Contact: Yoshinide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suenho-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>

, Carninci, P., Shibata, Y., Hayatsu, M., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

Wag1, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

e mouse tissues.

#### FEATURES

##### source

location/Qualifiers  
1. 479  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="G370022K07"  
/clone\_id="RIKEN full-length enriched, B16 F10Y cells"  
/cell\_type="B16 F10Y cells"

BASE COUNT 112 a 109 c 149 g 109 t  
ORIGIN

#### Query Match

65.5%; Score 19; DB 9; Length 479;

Best Local Similarity 81.5%; Pred. No. 1,1e+03; Mismatches 5; Indels 0; Gaps 0;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
OY 1 cagcgaaagcgctcgccagcagcagc 27  
|||||  
DB 45 CGCGGTAGCGCGCGCGCGCGCGCG 19

Search completed: July 31, 2002, 12:11:53  
Job time: 13466 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:11:05 ; Search time 720.8 Seconds

(without alignments)  
69.077 Million cell updates/sec

Title: us-09-899-718a-2

Perfect score: 29

Sequence: 1 caacgcaaaagcgctgcgcacgacacacac 29

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

N\_Geneseq\_032802:\*

1: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:\*

2: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:\*

3: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:\*

4: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:\*

5: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:\*

6: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:\*

7: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:\*

8: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:\*

9: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:\*

10: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:\*

11: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:\*

12: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:\*

13: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:\*

14: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:\*

15: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:\*

16: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:\*

17: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:\*

18: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:\*

19: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:\*

20: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:\*

21: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:\*

22: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:\*

23: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:\*

24: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| C 1        | 19.6  | 67.6        | 2473   | 22    | AAAL05200   |
| C 2        | 19.6  | 67.6        | 2474   | 22    | AAAL05201   |
| C 3        | 18.6  | 64.1        | 991    | 21    | AAAT4887    |
| 4          | 18.6  | 64.1        | 3546   | 16    | AAO90656    |
| 5          | 18.6  | 64.1        | 3591   | 16    | AAO90658    |
| 6          | 18.4  | 63.4        | 1511   | 22    | AAAF74776   |
| 7          | 18.4  | 63.4        | 3641   | 22    | AAAF74767   |
| 8          | 18.2  | 62.8        | 2557   | 23    | ABLI8098    |
| 9          | 18    | 62.1        | 3810   | 21    | AACT7273    |

|    |      |      |         |    |           |                      |
|----|------|------|---------|----|-----------|----------------------|
| 10 | 18   | 62.1 | 4022    | 21 | AAAC81226 | Human tyrosine pho   |
| 11 | 18   | 62.1 | 4456    | 20 | AAAB81748 | Human AIP encoding   |
| 12 | 18   | 62.1 | 5234    | 20 | AAAC81224 | Human histidine do   |
| 13 | 18   | 62.1 | 5436    | 21 | AAAC81227 | Human tyrosine pho   |
| 14 | 18   | 62.1 | 7016    | 22 | AAK67154  | Human immune/haema   |
| 15 | 18   | 62.1 | 7016    | 22 | AAK70464  | Human immune/haema   |
| 16 | 18   | 62.1 | 7016    | 22 | AAK74836  | Human immune/haema   |
| 17 | 18   | 62.1 | 9309    | 21 | AAAC81225 | Human tyrosine pho   |
| 18 | 18   | 62.1 | 10427   | 21 | AAZ36325  | Mechanical stress    |
| 19 | 18   | 62.1 | 14567   | 23 | AAAT70214 | Human encoding novel |
| 20 | 18   | 62.1 | 15214   | 22 | AAH98489  | Human EST-derived    |
| 21 | 18   | 62.1 | 16020   | 21 | AAA39283  | Streptomyces nogal   |
| 22 | 17.8 | 61.4 | 969     | 22 | AAAB68449 | C glutamicum codin   |
| 23 | 17.8 | 61.4 | 1092    | 22 | AAAF71371 | Corynebacterium gl   |
| 24 | 17.8 | 61.4 | 1540    | 22 | AAAT20978 | C glutamicum glk     |
| 25 | 17.8 | 61.4 | 34980   | 22 | AAH68531  | C glutamicum codin   |
| 26 | 17.4 | 60.0 | 636     | 23 | ABL21553  | Drosophila melanog   |
| 27 | 17.4 | 60.0 | 728     | 23 | AAAB81871 | Drosophila melanog   |
| 28 | 17.4 | 60.0 | 728     | 23 | AAAB85709 | DNA encoding novel   |
| 29 | 17.4 | 60.0 | 753     | 22 | ABAL6726  | Human nervous syst   |
| 30 | 17.4 | 60.0 | 1440    | 23 | AAAB81872 | DNA encoding novel   |
| 31 | 17.4 | 60.0 | 1440    | 23 | AAAB85699 | DNA encoding novel   |
| 32 | 17.4 | 60.0 | 1488    | 22 | AAAB81457 | Escherichia coli p   |
| 33 | 17.4 | 60.0 | 1488    | 23 | AAAS52292 | E. coli DNA for ce   |
| 34 | 17.4 | 60.0 | 2322    | 23 | AAAT98855 | DNA encoding novel   |
| 35 | 17.4 | 60.0 | 2556    | 23 | AAAT72099 | Drosophila melanog   |
| 36 | 17.4 | 60.0 | 2706    | 23 | ABL21552  | Drosophila melanog   |
| 37 | 17.4 | 60.0 | 2881    | 22 | AAAS34692 | Human DNA for a no   |
| 38 | 17.4 | 60.0 | 2888    | 22 | AAAS34691 | Human DNA for a no   |
| 39 | 17.4 | 60.0 | 2946    | 23 | AAAB82286 | DNA encoding novel   |
| 40 | 17.4 | 60.0 | 3054    | 22 | AAH17585  | Human cDNA sequenc   |
| 41 | 17.4 | 60.0 | 4037    | 23 | ABLI09561 | Drosophila melanog   |
| 42 | 17.4 | 60.0 | 4335    | 23 | AAAT98854 | DNA encoding novel   |
| 43 | 17.4 | 60.0 | 4891    | 23 | ABLI8392  | Drosophila melanog   |
| 44 | 17.4 | 60.0 | 12081   | 23 | ABLI09560 | Drosophila melanog   |
| 45 | 17.4 | 60.0 | 4403765 | 22 | AAI99683  | Mycobacterium tube   |

#### ALIGNMENTS

|          |                                                                           |
|----------|---------------------------------------------------------------------------|
| RESULT 1 | AAAL05200/c                                                               |
| ID       | AAAL05200 standard; DNA; 2473 BP.                                         |
| XX       | AAAL05200;                                                                |
| AC       | 21-NOV-2001 (first entry)                                                 |
| XX       | Human reproductive system related antigen DNA SEQ ID NO: 7888.            |
| DE       | Human reproductive system related antigen; reproductive system disorder;  |
| XX       | Human: reproductive system related antigen; reproductive system disorder; |
| KW       | cancer; gene therapy; ds.                                                 |
| XX       | Human sapiens.                                                            |
| OS       | WO200155320-A2.                                                           |
| XX       | 02-AUG-2001.                                                              |
| PD       | 17-JAN-2001; 2001WO-US01339.                                              |
| XX       | 31-JAN-2000; 2000US-0179065.                                              |
| PR       | 04-FEB-2000; 2000US-0180628.                                              |
| PR       | 24-FEB-2000; 2000US-0184664.                                              |
| PR       | 02-MAR-2000; 2000US-0186350.                                              |
| PR       | 16-MAR-2000; 2000US-0189874.                                              |
| PR       | 17-MAR-2000; 2000US-0190076.                                              |
| PR       | 18-APR-2000; 2000US-0198123.                                              |
| PR       | 19-MAY-2000; 2000US-0205151.                                              |
| PR       | 07-JUN-2000; 2000US-0209467.                                              |
| PR       | 28-JUN-2000; 2000US-0214886.                                              |
| PR       | 30-JUN-2000; 2000US-0215135.                                              |





Query Match 67.68; Score 19.6; DB 22; Length 2473;  
Best Local Similarity 84.68; Pred. No. 44;  
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 caccgaagcgcgtcgcgcagccac 26  
|||||  
Db 1678 CACGCAAGGCCCTCAGCTTGCAC 1653

RESULT 2  
AAL05201/c  
ID AAL05201 standard; DNA; 2474 BP.

AC AAL05201;  
DT 21-NOV-2001 (first entry)  
XX  
DE Human reproductive system related antigen DNA SEQ ID NO: 7889.  
KW Human; reproductive system related antigen; reproductive system disorder;  
KW cancer; gene therapy; ds.  
XX  
OS Homo sapiens.  
XX  
PN W0200155320-A2.  
PD  
XX 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01339.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184654.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227189.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241121.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246539.  
PR 08-NOV-2000; 2000US-0246602.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.

```

PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251088.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251859.
PR 08-DEC-2000; 2000US-0251859.
PR 08-DEC-2000; 2000US-0251859.
PR 08-DEC-2000; 2000US-0251859.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-46570/50.
DR
XX Isolated nucleic acid molecule encoding a reproductive system antigen -
PT is used in preventing, treating or ameliorating a medical condition -
XX
XX Disclosure; SEQ ID NO 7889; 1297bp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders.
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 2474 BP; 431 A; 863 C; 812 G; 368 T; 0 other;

```

```

Query Match 67.6%; Score 19.6; DB 22; Length 2474;
Best Local Similarity 84.6%; Pred. No. 44;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Oy 1 caccgaagcgctgcgcagccac 26
    |||||||
Db 1679 CACGCAAGGCCGCTGCTGCCAC 1654

```

```

RESULT 3
AAFL487/C
ID AAFL487 standard; cDNA; 991 BP.
AC AAFL487;
XX
XX 13-MAR-2001 (first entry)
XX
XX Trichoderma reesei EST SEQ ID NO:7410.
DE
XX Multiple gene expression; filamentous fungal cell; EST;
XX expressed sequence tag; Fusarium venenatum; Aspergillus niger;
XX Aspergillus oryzae; Trichoderma reesei; identification; recombination;
XX culture condition; environmental stress; spore morphogenesis;
XX metabolic pathway engineering; catabolic pathway engineering; ss.
XX
XX Trichoderma reesei.
OS
XX WO200056762-A2.
XX
XX 28-SEP-2000.
XX
XX 22-MAR-2000; 2000WO-US0781.
PF

```

```

XX
XX 22-MAR-1999; 99US-0273623.
XX
XX (NOVO) NOVO NORDISK BIOTECH INC.
XX (NOVO) NOVO NORDISK AS.
XX
XX Berka RM, Rey MW, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;
XX WPI; 2000-594572/56.
XX
XX Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
XX Claim 89; Page 2992; 3161pp; English.
XX
XX The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic advantages over genomic or
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organization of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Trichoderma reesei; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
XX

```

```

SQ Sequence 991 BP; 200 A; 290 C; 269 G; 220 T; 12 other;

```

```

Query Match 64.1%; Score 18.6; DB 21; Length 991;
Best Local Similarity 84.0%; Pred. No. 11e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Oy 1 caccgaagcgctgcgcagccca 25
    |||||||
Db 621 CACGCAAGGCCGCTGCCAGTCA 597

```

```

RESULT 4
AAQ90656
ID AAQ90656 standard; cDNA; 3546 BP.
AC AAQ90656;
XX
XX 11-NOV-1995 (first entry)
XX
XX Eph-related PTK Cdk10 cDNA.
DE
XX Cdk10; Eph; protein tyrosine-kinase; PTK; cancer; diagnosis;
XX prognosis; ss.
XX
XX Gallus sp.
OS
XX
XX Key Location/Qualifiers
XX CDS 2..2923
XX FT /*tag= a
XX
XX WO9515375-A.
PN

```

XX 08-JUN-1995.  
 PD  
 XX  
 PF 07-SEP-1994; 94WO-US10140.  
 XX  
 PR 03-DEC-1993; 93US-0162809.  
 XX  
 PA (LJOL-) LA JOLLA CANCER RES FOUND.  
 XX  
 PI Pasquale EB, Sajjadi FG;  
 XX  
 DR WPI: 1995-215256/28.  
 XX  
 P-PSDB; AAR75708.  
 XX  
 PT Eph-related protein tyrosine kinase(s) - for monitoring and diagnosing  
 PT cancer.  
 XX  
 PS Claim 2; Page 63-67; 129pp; English.  
 XX  
 CC Novel Eph-related PTK cDNA clone Cek10 (given in AAQ90656) and a  
 CC variant clone, Cek10+ (AAQ90658), whose product contains a 15-amino  
 CC acid insertion in the juxtamembrane domain, were isolated from a  
 CC chick embryo library in lambda gt11. Cek10 and Cek10+ may  
 CC originate from the same gene by alternative splicing. Cek10  
 CC expression was prominent in the kidney, and to a lesser extent in  
 CC the lung.  
 CC  
 SQ Sequence 3546 BP; 742 A; 1100 C; 1031 G; 673 T; 0 other;

Query Match 64.1%; Score 18.6; DB 16; Length 3546;  
 Best Local Similarity 84.0%; Pred. No. 1.2e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 OY 3 cgcgaagcggtgcgcagccagc 27  
 ||||| | ||||| ||||| |||||  
 Db 2363 cgcgaattcagctgcgcagccagc 2387

RESULT 5  
 AAQ90658  
 ID AAQ90658 standard; cDNA; 3591 BP.  
 XX  
 AC AAQ90658;  
 XX  
 DT 11-NOV-1995 (first entry)  
 XX  
 DE Eph-related PTK Cek10+ cDNA.  
 XX  
 KW Cek10+; Eph; protein tyrosine-kinase; PTK; cancer; diagnosis;  
 KW prognosis; ss.  
 XX  
 OS Gallus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 2..2968  
 FT /\*tag= a  
 FT  
 PN WO9515375-A.  
 XX  
 PD 08-JUN-1995.  
 XX  
 PF 07-SEP-1994; 94WO-US10140.  
 XX  
 PR 03-DEC-1993; 93US-0162809.  
 XX  
 PA (LJOL-) LA JOLLA CANCER RES FOUND.  
 XX  
 PI Pasquale EB, Sajjadi FG;  
 XX  
 DR WPI: 1995-215256/28.  
 DR P-PSDB; AAR75710.  
 XX

PT Eph-related protein tyrosine kinase(s) - for monitoring and diagnosing  
 PT cancer.  
 XX  
 PS Claim 2; Page 78-82; 129pp; English.  
 XX  
 CC Novel Eph-related PTK cDNA clone Cek10 (given in AAQ90656) and a  
 CC variant clone, Cek10+ (AAQ90658), whose product contains a 15-amino  
 CC acid insertion in the juxtamembrane domain, were isolated from a  
 CC chick embryo library in lambda gt11. Cek10 and Cek10+ may  
 CC originate from the same gene by alternative splicing. Cek10  
 CC expression was prominent in the kidney, and to a lesser extent in  
 CC the lung.  
 CC  
 SQ Sequence 3591 BP; 753 A; 1112 C; 1042 G; 684 T; 0 other;

Query Match 64.1%; Score 18.6; DB 16; Length 3591;  
 Best Local Similarity 84.0%; Pred. No. 1.2e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 OY 3 cgcgaagcggtgcgcagccagc 27  
 ||||| | ||||| ||||| |||||  
 Db 2408 cgcgaattcagctgcgcagccagc 2432

RESULT 6  
 AAF74776  
 ID AAF74776 standard; cDNA; 1511 BP.  
 XX  
 AC AAF74776;  
 XX  
 DT 17-MAY-2001 (first entry)  
 XX  
 DE Human DEC2b encoding cDNA sequence SEQ ID NO:11.  
 XX  
 KW DEC2a; DEC2b; bHLH type transcription factor; DEC2; DEC1;  
 KW basic helix loop helix protein; cell differentiation; proliferation; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 2..1456  
 FT /\*tag= a  
 FT /product= "DEC2b"  
 FT /note= "bHLH type transcription factor"  
 XX  
 PN WO200114551-A1.  
 XX  
 PD 01-MAR-2001.  
 XX  
 PF 19-JUN-2000; 2000WO-JP03991.  
 XX  
 PR 19-AUG-1999; 99JP-0233286.  
 XX  
 PA (CHUS ) CHUGAI SEIYAKU KK.  
 XX  
 PI Fujimoto K, Shin M, Kato Y;  
 XX  
 DR WPI: 2001-202935/20.  
 DR P-PSDB; AAB70693.  
 XX  
 PT DEC2 is a basic helix loop helix protein of the DEC family for use in  
 PT development of drugs for treatment of disorders of cell differentiation  
 PT and proliferation -  
 XX  
 PS Claim 1; Page 63-68; 83pp; Japanese.  
 XX  
 CC The present invention describes a basic helix loop helix (bHLH) type  
 CC transcription factor designated DEC2. DEC2 can be used as a tool in  
 CC the development of drugs for the treatment and prevention of disorders  
 CC involving cell differentiation and proliferation. The present sequence  
 CC encodes the specifically claimed human DEC2b protein, as given in the  
 CC present invention.

XX Sequence 1511 BP; 302 A; 521 C; 449 G; 239 T; 0 other;  
SQ

Query Match 63.4%; Score 18.4; DB 22; Length 1511;  
Best Local Similarity 78.6%; Pred. No. 1.4e+02;  
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 accgaagcgctgcgcgcacacacac 29  
| | | | | | | | | | | | | | | | | |  
DB 769 agcgaagcgcgcgcgagcgacgcgctc 736

## RESULT 7

AAE74767  
ID AAE74767 standard; cDNA; 3641 BP.

AC AAE74767;

DT 17-MAY-2001 (first entry)

DE Human DEC2a encoding CDNA sequence SEQ ID NO:1.

KW DEC2a; DEC2b; bHLH type transcription factor; DEC2; DEC1;

KW basic helix loop helix protein; cell differentiation; proliferation; ss.

OS Homo sapiens.

PH Key Location/Qualifiers

FT CDS 135..1583

FT /\*tag= a  
/product= "DEC2a"  
/note= "bHLH type transcription factor"

PN MO200114551-A1.

PD 01-MAR-2001.

PF 19-JUN-2000; 2000MO-JP03991.

PR 19-AUG-1999; 99JP-0233286.

PA (CHUS ) CHUGAI SEIYAKU KK.

PI Fujimoto K, Shin M, Kato Y;

DR WPI: 2001-202935/20.

DR P-PSDB; AAE70692.

XX DEC2 is a basic helix loop helix protein of the DEC family for use in

PT development of drugs for treatment of disorders of cell differentiation

PT and proliferation -

PS Claim 1: Page 48-55; 83pp; Japanese.

XX The present invention describes a basic helix loop helix (bHLH) type

CC transcription factor designated DEC2. DEC2 can be used as a tool in

CC the development of drugs for the treatment and prevention of disorders

CC involving cell differentiation and proliferation. The present sequence

CC encodes the specifically claimed human DEC2a protein, as given in the

CC present invention.

XX Sequence 3641 BP; 994 A; 922 C; 836 G; 889 T; 0 other;

Query Match 63.4%; Score 18.4; DB 22; Length 3641;

Best Local Similarity 78.6%; Pred. No. 1.4e+02;  
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 accgaagcgctgcgcgcacacacac 29

DB 896 agcgaagcgcgcgagcgacgcgctc 923

## RESULT 8

ABLI8098  
ID ABLI8098 standard; DNA; 2557 BP.

AC ABLI8098;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 5767.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI: 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

PS Claim 1: SEQ ID NO 5767; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA

CC sequences (ABLI01840-ABLI6175) and the encoded proteins

CC (ABBS7737-ABBS7072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 2557 BP; 671 A; 546 C; 563 G; 777 T; 0 other;

Query Match 62.8%; Score 18.2; DB 23; Length 2557;  
Best Local Similarity 87.0%; Pred. No. 1.7e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 cgcgaagcgcgctgcgcgcacacac 25

DB 1779 cccacagcgcgctgcgcgcacacac 1801

## RESULT 9

AACT7273  
ID AACT7273 standard; cDNA; 3810 BP.

AC AACT7273;

DT 08-FEB-2001 (first entry)

DE Human ORFX ORF2828 polynucleotide sequence SEQ ID NO:5655.

XX Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;  
XX vulnerable; antiproliferative; antiparkinsonian; neuroprotective;



```

XX DE Human ALP encoding cDNA.
XX AC
XX KW PRP04; PRP05; PRP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
XX KW type I receptor serine/threonine kinase; cancer; leukemia; lymphoma;
XX KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
XX KW Parkinson's disease; Huntington's disease; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 313..4137
XX FT /tag= a
XX PN WO9849317-A2.
XX PD 05-NOV-1998.
XX PE 27-APR-1998; 98WO-US08439.
XX PR 23-OCT-1997; 97US-0063595.
XX PR 28-APR-1997; 97US-0044428.
XX PR 20-MAY-1997; 97US-0047222.
XX PR 11-JUN-1997; 97US-0049477.
XX PR 11-JUN-1997; 97US-0049756.
XX PR 18-JUN-1997; 97US-0049914.
XX PA (SUGEN-) SUGEN INC.
XX PI App H, Clary D, Courtneidge SA, Hui TH, Jallat B;
XX PI Markby D, Onrust S, Peles E, Plowman GD;
XX PI WPI: 1999-009434/01.
XX DR P-PSDB; AAW89253.
XX PT New nucleic acid encoding specific protein tyrosine phosphatases -
XX PT useful for identifying specific modulators for treatment and
XX PT prevention of cancer and neurodegenerative disease
XX PS Claim 2; Page 149-150; 199pp; English.
XX XX The present invention describes isolated, enriched or purified nucleic
XX CC acids encoding PRP04, SAD, PRP05, PRP10, ALP and ALK-7 proteins. The
XX CC present sequence encodes human ALP. The above proteins, other than
XX CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
XX CC substances that modulate their activity (i.e. agonists and antagonists,
XX CC including NBP) in vivo or in vitro. These substances are used to treat
XX CC or prevent diseases associated with abnormal signal transduction
XX CC pathways that involve the proteins, particularly cancer (e.g. leukemia
XX CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
XX CC serine/threonine kinase) are used to promote neuronal survival.
XX CC particularly for treating Alzheimer's, Parkinson's or Huntington's
XX CC diseases. Nucleic acid fragments of the polynucleotides encoding the
XX CC proteins can be used as probes to identify and clone related sequences;
XX CC to detect protein-encoded RNA; to generate transgenic animals and in
XX CC gene therapy (optionally after mutation). Ab are used to determine the
XX CC proteins.
XX SQ Sequence 4456 BP; 929 A; 1564 C; 1184 G; 779 T; 0 other:
OY Query Match 62.1%; Score 18; DB 20; Length 4456;
Best Local Similarity 80.8%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 4 gcaaaagcgctgcgcagcagcagc 29
DB 1846 gccatgcyggttcgcccagccaccac 1871
RESULT 12
AAC81224
ID AAC81224 standard; cDNA; 5234 BP.

```

```

XX AC AAC81224;
XX XX 23-FEB-2001 (first entry)
XX DT
XX DE Human histidine domain-protein tyrosine phosphatase cDNA, SEQ ID NO:1.
XX KW Human; histidine domain-protein tyrosine phosphatase; HD-PTP;
XX KW chromosome 3p21.3; gene deletion; tumour suppressor; cytostatic;
XX KW lung cancer; tumour; gene therapy; diagnosis; recombinant production;
XX KW anticancer; ss.
XX OS Homo sapiens.
XX FH WO200063392-A1.
XX FT 26-OCT-2000.
XX PD 14-APR-2000; 2000WO-JP02455.
XX PE 16-APR-1999; 99JP-010842.
XX PR (KYOW ) KYOWA HAKKO KOGYO KK.
XX PA Shintzu K;
XX PI WPI: 2000-672740/65.
XX DR P-PSDB; AAB29661.
XX PT Human tyrosine phosphatase with oncostatic activity encoded by a gene
XX PT frequently deleted in lung cancer, useful for treatment and diagnosis
XX PT of tumors
XX PS Claim 4; Page 58-69; 134pp; Japanese.
XX XX The invention relates to a novel human tyrosine phosphatase, histidine
XX CC domain-protein tyrosine phosphatase (HD-PTP; AAB29661) and to human
XX CC HD-PTP nucleic acids (AAC81224, AAC81225, AAC81263). The
XX CC HD-PTP gene is located on chromosome 3p21.3. This region is frequently
XX CC found to be deleted in lung cancers, and is therefore thought to contain
XX CC a tumour suppressor gene. The invention also relates to expression
XX CC vectors and host cells containing human HD-PTP nucleic acids; the
XX CC recombinant production of HD-PTP; anticancer drugs containing HD-PTP;
XX CC gene therapy compositions containing DNA encoding HD-PTP; diagnostic
XX CC reagents containing HD-PTP oligonucleotides; antibodies specific for
XX CC HD-PTP; and an immunoassay method using HD-PTP-specific antibodies for
XX CC use in cancer diagnosis and investigation. HD-PTP proteins, nucleic acids
XX CC and antibodies may be used in the treatment, investigation and diagnosis
XX CC of cancers, particularly those of the lung. The present sequence
XX CC represents cDNA encoding human HD-PTP.
XX SQ Sequence 5234 BP; 1049 A; 1803 C; 1441 G; 941 T; 0 other:
OY Query Match 62.1%; Score 18; DB 21; Length 5234;
Best Local Similarity 80.8%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 4 gcaaaagcgctgcgcagcagcagc 29
DB 2683 gccatgcyggttcgcccagccaccac 2708
RESULT 13
AAC81227
ID AAC81227 standard; DNA; 5436 BP.
XX AC AAC81227;
XX XX 23-FEB-2001 (first entry)
XX DT
XX DE Human tyrosine phosphatase HD-PTP DNA fragment cKAL11, SEQ ID NO:5.
XX XX

```



|    |                                 |                |
|----|---------------------------------|----------------|
| PR | 29-SEP-2000                     | 2000US-0236368 |
| PR | 29-SEP-2000                     | 2000US-0236369 |
| PR | 29-SEP-2000                     | 2000US-0236370 |
| PR | 02-OCT-2000                     | 2000US-0236802 |
| PR | 02-OCT-2000                     | 2000US-0237037 |
| PR | 02-OCT-2000                     | 2000US-0237038 |
| PR | 02-OCT-2000                     | 2000US-0237039 |
| PR | 02-OCT-2000                     | 2000US-0237040 |
| PR | 13-OCT-2000                     | 2000US-0239935 |
| PR | 13-OCT-2000                     | 2000US-0239937 |
| PR | 20-OCT-2000                     | 2000US-0241221 |
| PR | 20-OCT-2000                     | 2000US-0241221 |
| PR | 20-OCT-2000                     | 2000US-0241785 |
| PR | 20-OCT-2000                     | 2000US-0241786 |
| PR | 20-OCT-2000                     | 2000US-0241787 |
| PR | 08-NOV-2000                     | 2000US-0246475 |
| PR | 08-NOV-2000                     | 2000US-0246476 |
| PR | 08-NOV-2000                     | 2000US-0246477 |
| PR | 08-NOV-2000                     | 2000US-0246523 |
| PR | 08-NOV-2000                     | 2000US-0246524 |
| PR | 08-NOV-2000                     | 2000US-0246525 |
| PR | 08-NOV-2000                     | 2000US-0246526 |
| PR | 08-NOV-2000                     | 2000US-0246527 |
| PR | 08-NOV-2000                     | 2000US-0246528 |
| PR | 08-NOV-2000                     | 2000US-0246532 |
| PR | 08-NOV-2000                     | 2000US-0246539 |
| PR | 08-NOV-2000                     | 2000US-0246610 |
| PR | 08-NOV-2000                     | 2000US-0246611 |
| PR | 17-NOV-2000                     | 2000US-0249207 |
| PR | 17-NOV-2000                     | 2000US-0249208 |
| PR | 17-NOV-2000                     | 2000US-0249209 |
| PR | 17-NOV-2000                     | 2000US-0249210 |
| PR | 17-NOV-2000                     | 2000US-0249211 |
| PR | 17-NOV-2000                     | 2000US-0249212 |
| PR | 17-NOV-2000                     | 2000US-0249213 |
| PR | 17-NOV-2000                     | 2000US-0249214 |
| PR | 17-NOV-2000                     | 2000US-0249215 |
| PR | 17-NOV-2000                     | 2000US-0249216 |
| PR | 17-NOV-2000                     | 2000US-0249217 |
| PR | 17-NOV-2000                     | 2000US-0249218 |
| PR | 17-NOV-2000                     | 2000US-0249244 |
| PR | 17-NOV-2000                     | 2000US-0249245 |
| PR | 17-NOV-2000                     | 2000US-0249264 |
| PR | 17-NOV-2000                     | 2000US-0249265 |
| PR | 17-NOV-2000                     | 2000US-0249297 |
| PR | 17-NOV-2000                     | 2000US-0249299 |
| PR | 17-NOV-2000                     | 2000US-0249300 |
| PR | 01-DEC-2000                     | 2000US-0250160 |
| PR | 01-DEC-2000                     | 2000US-0250391 |
| PR | 05-DEC-2000                     | 2000US-0251030 |
| PR | 05-DEC-2000                     | 2000US-0251038 |
| PR | 05-DEC-2000                     | 2000US-0251679 |
| PR | 06-DEC-2000                     | 2000US-0251479 |
| PR | 08-DEC-2000                     | 2000US-0251856 |
| PR | 08-DEC-2000                     | 2000US-0251868 |
| PR | 08-DEC-2000                     | 2000US-0251869 |
| PR | 08-DEC-2000                     | 2000US-0251989 |
| PR | 08-DEC-2000                     | 2000US-0251990 |
| PR | 11-DEC-2000                     | 2000US-0254097 |
| PR | 05-JAN-2001                     | 2000US-0259678 |
| XX |                                 |                |
| XX |                                 |                |
| PA | (HUMA - ) HUMAN GENOME SCI INC. |                |
| XX |                                 |                |
| PI | Rosen CA, Barash SC, Ruben SM,  |                |
| DR | WPI, 2001-483426/52.            |                |

|           |                                                                           |
|-----------|---------------------------------------------------------------------------|
| XX        | Nucleic acids encoding human immune/hematopoietic antigen polypeptides,   |
| PT        | useful for preventing, diagnosing and/or treating cancers and             |
| PT        | metastasis -                                                              |
| PS        | -                                                                         |
| XX        | Disclosure; SEQ ID NO 21966; 3071bp + Sequence Listing; English.          |
| CC        | AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)    |
| CC        | antino acid sequences given in AAK82170 to AAM91921. (I) have cytostatic  |
| CC        | activity, and can be used in gene therapy and vaccine production. (I)     |
| CC        | proteins and polynucleotides may be used in the prevention, diagnosis and |
| CC        | treatment of diseases associated with inappropriate (I) expression. For   |
| CC        | example, they may be used to treat disorders associated with decreased    |
| CC        | expression by rectifying mutations or deletions in a patient's genome     |
| CC        | that affect the activity of (I) by expressing inactive proteins or to     |
| CC        | supplement the patients own production of (I). Additionally, (I)          |
| CC        | polynucleotides may be used to produce the secreted (I), by inserting the |
| CC        | nucleic acids into a host cell and culturing the cell to express the      |
| CC        | protein. (I) proteins and polynucleotides may be used to prevent,         |
| CC        | diagnose and treat immune/hematopoietic-related diseases, especially      |
| CC        | cancers and cancer metastases of haematopoietic-derived cells. AAK64703   |
| CC        | to AAK87694 represent human immune/hematopoietic antigen genomic          |
| CC        | sequences from the present invention. AAK54942 to AAK54950 and AAM2169    |
| CC        | represent sequences used in the exemplification of the present invention. |
| XX        | Sequence 7016 BP; 1339 A; 2079 C; 2321 G; 1277 T; 0 other;                |
| OY        | 4 gcaaaagcgcgtcgccacgaccagac 29                                           |
| DB        | 3019 GCCATGCGCGTTCGCCACGCCACCAC 2994                                      |
| RESULT 15 |                                                                           |
| ID        | AAK70464/c                                                                |
| XX        | AAK70464 standard; DNA; 7016 BP.                                          |
| AC        | AAK70464;                                                                 |
| DT        | 06-NOV-2001 (first entry)                                                 |
| DE        | Human immune/hematopoietic antigen genomic sequence SEQ ID NO:25276.      |
| KM        | Human; immune; haematopoietic; immune/hematopoietic antigen; cancer;      |
| KW        | cytostatic; gene therapy; vaccine; metastasis; ds.                        |
| OS        | Homo sapiens.                                                             |
| PN        | WO200157182-A2.                                                           |
| PD        | 09-AUG-2001.                                                              |
| PF        | 17-JAN-2001; 200IWO-US01354.                                              |
| PR        | 31-JAN-2000; 200OUS-0179065.                                              |
| PR        | 04-FEB-2000; 200OUS-0180628.                                              |
| PR        | 24-FEB-2000; 200OUS-0184664.                                              |
| PR        | 02-MAR-2000; 200OUS-0186350.                                              |
| PR        | 16-MAR-2000; 200OUS-0189874.                                              |
| PR        | 17-MAR-2000; 200OUS-0190076.                                              |
| PR        | 18-APR-2000; 200OUS-0198123.                                              |
| PR        | 19-MAY-2000; 200OUS-0205515.                                              |
| PR        | 07-JUN-2000; 200OUS-0209467.                                              |
| PR        | 28-JUN-2000; 200OUS-0214886.                                              |
| PR        | 30-JUN-2000; 200OUS-0215135.                                              |
| PR        | 07-JUL-2000; 200OUS-0216647.                                              |
| PR        | 07-JUL-2000; 200OUS-0216880.                                              |
| PR        | 11-JUL-2000; 200OUS-0217487.                                              |
| PR        | 11-JUL-2000; 200OUS-0217496.                                              |





CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AA64703  
CC to AA687694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AA55942 to AA55950 and AA62169  
CC represent sequences used in the exemplification of the present invention.  
XX  
XX Sequence 7016 BP, 1339 A; 2079 C; 2321 G; 1277 T; 0 other;  
XX

|                          |        |                    |           |              |
|--------------------------|--------|--------------------|-----------|--------------|
| Query Match              | 62.18; | Score 18;          | DB 22;    | Length 7016; |
| Best Local Similarity    | 80.8%; | Pred. No. 2.1e+02; |           |              |
| Matches 21; Conservative | 0;     | Mismatches 5;      | Indels 0; | Gaps 0;      |

Qy 4 gcaagcgcgtcgccagccacgac 29  
||| ||||| ||||| |||  
Db 3019 GCCATGGCGGTTCCGCCAGCCACCAC 2994

Search completed: July 31, 2002, 14:11:09  
Job time: 17366 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:00:36 ; Search time 6034.22 Seconds

(Without alignments)  
100.571 Million cell updates/sec

Title: US-09-899-718A-2

Perfect score: 29

Sequence: 1 caccgaagcgcgctgcgcagcagcagc 29

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 100%

Listing first 45 summaries

Database :  
1: gb\_ba:\*  
2: gb\_hgt:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hgt\_hum:\*  
31: em\_hgt\_inv:\*  
32: em\_hgt\_other:\*  
33: em\_hgtgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result Query  
No. Score Match Length DB ID Description

|    |      |       |        |    |           |                      |
|----|------|-------|--------|----|-----------|----------------------|
| 1  | 29   | 100.0 | 29     | 6  | AX349064  | AX349064 Sequence    |
| 2  | 29   | 100.0 | 3785   | 6  | AX349063  | AX349063 Sequence    |
| 3  | 21   | 72.4  | 1437   | 14 | NPOG37DP  | D1306 Oryza pseu     |
| 4  | 21   | 72.4  | 5388   | 14 | OPU39145  | U39145 Oryza pseu    |
| 5  | 21   | 72.4  | 131995 | 14 | OPU75930  | U75930 Oryza pseu    |
| 6  | 20.2 | 69.7  | 35654  | 1  | SC7H1     | AL021411 Streptomy   |
| 7  | 20   | 69.0  | 162623 | 2  | AC106098  | AC106098 Rattus no   |
| 8  | 19.8 | 68.3  | 2053   | 9  | BC011138  | BC011138 Homo sapi   |
| 9  | 19.8 | 68.3  | 4926   | 1  | AF357175  | AF357175 Mycobacte   |
| 10 | 19.8 | 68.3  | 5047   | 9  | AB011161  | AB011161 Homo sapi   |
| 11 | 19.8 | 68.3  | 33818  | 1  | MTCY78    | Z77165 Mycobacteri   |
| 12 | 19.8 | 68.3  | 38209  | 2  | AC005542  | AC005542 Homo sapi   |
| 13 | 19.8 | 68.3  | 175625 | 2  | AC068475  | AC068475 Homo sapi   |
| 14 | 19.6 | 67.6  | 15470  | 1  | AE004513  | AE004513 Pseudomon   |
| 15 | 19.6 | 67.6  | 26591  | 1  | AE008794  | AE008794 Salmoneil   |
| 16 | 19.6 | 67.6  | 34108  | 1  | SC25      | AL354048 Streptomy   |
| 17 | 19.6 | 67.6  | 159324 | 2  | AC027307  | AC027307 Homo sapi   |
| 18 | 19.6 | 67.6  | 250050 | 1  | AL627273  | AL627273 Salmoneil   |
| 19 | 19.4 | 66.9  | 190050 | 1  | AL646080  | AL646080 Ralstonia   |
| 20 | 19.2 | 66.2  | 45205  | 3  | UA1014    | UA1014 Caenorhabdi   |
| 21 | 19.2 | 66.2  | 104326 | 1  | AB070940  | AB070940 Streptomy   |
| 22 | 19   | 65.5  | 4876   | 8  | NEO08     | L40806 Neurospora    |
| 23 | 19   | 65.5  | 10173  | 1  | AE008067  | AE008067 Agrobacte   |
| 24 | 19   | 65.5  | 10173  | 1  | AE009102  | AE009102 Agrobacte   |
| 25 | 19   | 65.5  | 38634  | 1  | SC2655    | AL359152 Streptomy   |
| 26 | 19   | 65.5  | 162566 | 2  | AL663032  | AL663032 Mus muscu   |
| 27 | 18.6 | 64.1  | 3324   | 2  | RCE245615 | AL243615 Rudriyiva   |
| 28 | 18.6 | 64.1  | 3546   | 6  | 115004    | 115004 Sequence 9    |
| 29 | 18.6 | 64.1  | 3591   | 5  | GGCEK10A  | 219061 G.gallus Ce   |
| 30 | 18.6 | 64.1  | 4180   | 1  | AF232751  | AF232751 Sequence 13 |
| 31 | 18.6 | 64.1  | 69304  | 2  | AC016497  | AC016497 Homo sapi   |
| 32 | 18.6 | 64.1  | 82909  | 1  | ECUDK367  | AC073867 Oryza sat   |
| 33 | 18.6 | 64.1  | 785    | 1  | ECUDK3    | X71492 E.coli udk    |
| 34 | 18.4 | 63.4  | 1511   | 6  | BD004347  | BD004347 Novel bHL   |
| 35 | 18.4 | 63.4  | 3641   | 9  | AB044088  | AB044088 Homo sapi   |
| 36 | 18.4 | 63.4  | 11267  | 1  | AE009589  | AE009589 Brucella    |
| 37 | 18.4 | 63.4  | 11267  | 1  | AE009098  | AE009098 Agrobacte   |
| 38 | 18.4 | 63.4  | 11267  | 1  | AE008063  | AE008063 Agrobacte   |
| 39 | 18.4 | 63.4  | 12996  | 1  | AE000296  | AE000296 Escherich   |
| 40 | 18.4 | 63.4  | 13252  | 1  | AE009527  | AE009527 Brucella    |
| 41 | 18.4 | 63.4  | 13430  | 1  | D90844    | D90844 E.coli geno   |
| 42 | 18.4 | 63.4  | 13918  | 1  | D90845    | D90845 E.coli geno   |
| 43 | 18.4 | 63.4  | 19366  | 1  | MCU15187  | U15187 Mycobacteri   |
| 44 | 18.4 | 63.4  | 36138  | 1  |           |                      |
| 45 | 18.4 | 63.4  |        |    |           |                      |

## ALIGNMENTS

|                                 |                                                             |                                   |     |        |                 |
|---------------------------------|-------------------------------------------------------------|-----------------------------------|-----|--------|-----------------|
| RESULT 1                        | AX349064                                                    | 29 bp                             | DNA | linear | PAT 06-FEB-2002 |
| LOCUS                           | AX349064                                                    | Sequence 2 from Patent WO0202785. |     |        |                 |
| DEFINITION                      | AX349064                                                    |                                   |     |        |                 |
| ACCESSION                       | AX349064                                                    |                                   |     |        |                 |
| VERSION                         | AX349064.1                                                  | GI:18615099                       |     |        |                 |
| KEYWORDS                        |                                                             |                                   |     |        |                 |
| SOURCE                          |                                                             |                                   |     |        |                 |
| ORGANISM                        |                                                             |                                   |     |        |                 |
| synthetic construct.            |                                                             |                                   |     |        |                 |
| artificial sequence.            |                                                             |                                   |     |        |                 |
| REFERENCE                       |                                                             |                                   |     |        |                 |
| 1 (sites)                       |                                                             |                                   |     |        |                 |
| AUTHORS                         | Sprunck,S., Kluth,A., Becker,D., Luetticke,S. and Loertz,H. |                                   |     |        |                 |
| TITLE                           | Promoters of gene expression in plant caryopses             |                                   |     |        |                 |
| JOURNAL                         | Patent: WO 0202785-A 2 10-JAN-2002;                         |                                   |     |        |                 |
| Avantis CropScience GmbH (DE)   |                                                             |                                   |     |        |                 |
| Location/Qualifiers             |                                                             |                                   |     |        |                 |
| 1..29                           |                                                             |                                   |     |        |                 |
| /organism="synthetic construct" |                                                             |                                   |     |        |                 |
| /db_xref="taxon:32630"          |                                                             |                                   |     |        |                 |
| /note="Oligonucleotide"         |                                                             |                                   |     |        |                 |
| BASE COUNT                      | 7 a                                                         | 12 c                              | 9 g | 1 t    |                 |
| ORIGIN                          |                                                             |                                   |     |        |                 |

Query Match 100.0%; Score 29; DB 6; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 0.84;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 caccgaagcgctgcgcagccagcagc 29  
 Db 1 CACGCAAGGCGCTCGCCAGCCAGCAGC 29

RESULT 2  
 LOCUS AX349063 3785 bp DNA linear PAT 06-FEB-2002  
 DEFINITION Sequence 1 from Patent WO0202785.  
 ACCESSION AX349063  
 VERSION AX349063.1 GI:18615098  
 KEYWORDS  
 SOURCE bread wheat.  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Pooidae; Triticeae; Triticum.  
 1 (sites)  
 Sprunck,S., Kluth,A., Becker,D., Luetticke,S. and Loerz,H.  
 Promoters of gene expression in plant caryopses  
 Patent: WO 0202785-A 1 10-JAN-2002;  
 Aventis CropScience GmbH (DE)  
 Location/Qualifiers  
 1..3785  
 /organism="Triticum aestivum"  
 /db\_xref="taxon:4565"

BASE COUNT 945 a 980 c 899 g 961 t  
 ORIGIN

Query Match 100.0%; Score 29; DB 6; Length 3785;  
 Best Local Similarity 100.0%; Pred. No. 0.34;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 caccgaagcgctgcgcagccagcagc 29  
 Db 1736 CACGCAAGGCGCTCGCCAGCCAGCAGC 1764

RESULT 3  
 LOCUS NP06P37DP 1437 bp DNA linear VRL 03-FEB-1999  
 DEFINITION Orygia pseudotsugata nuclear polyhedrosis virus genes for DNA  
 polymerase and spheroidin-like protein.  
 ACCESSION D13306.1 GI:222200  
 VERSION D13306  
 KEYWORDS DNA polymerase; spheroidin.  
 SOURCE Orygia pseudotsugata nuclear polyhedrosis virus DNA.  
 ORGANISM Orygia pseudotsugata single capsid nuclear polyhedrosis virus  
 Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
 Nucleopolyhedrovirus.  
 1 (bases 1 to 1437)  
 Gross,C.H., Wolgamot,G.M., Russell,R.L., Pearson,M.N. and  
 Rohmann,G.F.  
 A 37-kilodalton glycoprotein from a baculovirus of Orygia  
 pseudotsugata is localized to cytoplasmic inclusion bodies  
 J. Virol. 67 (1), 469-475 (1993)  
 93100831  
 COMMENT Submitted (28-SEP-1992) to DDBJ by:  
 George F. Rohmann  
 Dept. of Agricultural Chemistry  
 Oregon State University  
 Corvallis, Oregon 97331-6502  
 U.S.A.  
 Phone: 503-737-1793  
 Email: rohmann@crb.orst.edu  
 Fax: 503-737-0497.  
 Location/Qualifiers

source  
 1..1437  
 /organism="Orygia pseudotsugata single capsid nuclear  
 polyhedrosis virus"  
 /db\_xref="taxon:10450"  
 1..377  
 /gene="DNA polymerase"  
 <1..377  
 /gene="DNA polymerase"  
 /codon\_start=3  
 /product="DNA polymerase"  
 /protein\_id="BA02565.1"  
 /db\_xref="GI:222201"  
 /translation="AYPLRLEDAETMRISWKHTIINTPNMELLEITGDKNALAE  
 CYSALIEKYMHOQADKKRALVYKATRKAPASDASGRARAKAAPSDDESSSD  
 EDAPCEPKANNTFKCLYKAQ"  
 405..409  
 /note="late promoter"  
 421..1386  
 /gene="gp37"  
 421..1386  
 /gene="gp37"  
 /codon\_start=1  
 /product="gp37 (spheroidin-like protein)"  
 /protein\_id="BA02566.1"  
 /db\_xref="GI:222202"  
 /translation="MYKLCVAFALAVPAVBPBGYSTPYARQKCFADGNFYPWDNG  
 DGVDDEACRNAYKKEVFRYRAVGAPEGAALAAQYMFQOYAEYAAVAGPNRDELVK  
 REVLPHILGCAANDRHALFGKSGMDEPFNMRPDVLYVNRORASHFNVECPYAV  
 HEPSEFYVYTKFTNDRSRPWNLEFVIGGSGGLVPNGDAFCASGLVSIPIYVP  
 YRPGFVYVWQRIDPVEGFNCADIVFTENDEBRRYRAAKAVRODLNQNLND  
 CVEGPESCAPTRPQRRANHLIRGAGHDQADGASVRETIDEL"

BASE COUNT 300 a 462 c 444 g 231 t  
 ORIGIN

Query Match 72.4%; Score 21; DB 14; Length 1437;  
 Best Local Similarity 82.8%; Pred. No. 4.6e+02;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 caccgaagcgctgcgcagccagcagc 29  
 Db 259 CGCGCAAGGCGCGCCAGCCAGCAGC 287

RESULT 4  
 LOCUS OP039145/c 5388 bp DNA linear VRL 18-JUL-1996  
 DEFINITION Orygia pseudotsugata nuclear polyhedrosis virus DNA polymerase  
 gene, complete cds, and ORF 66 gene, partial cds.  
 ACCESSION U39145  
 VERSION U39145  
 KEYWORDS  
 SOURCE Orygia pseudotsugata nuclear polyhedrosis virus.  
 ORGANISM Orygia pseudotsugata single capsid nuclear polyhedrosis virus  
 Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
 Nucleopolyhedrovirus.  
 1 (bases 1 to 5388)  
 Gross,C.H., Wolgamot,G.M., Russell,R.L., Pearson,M.N. and  
 Rohmann,G.F.  
 A 37-kilodalton glycoprotein from a baculovirus of Orygia  
 pseudotsugata is localized to cytoplasmic inclusion bodies  
 J. Virol. 67 (1), 469-475 (1993)  
 93100831  
 COMMENT 2 (bases 1 to 5388)  
 Ahrens,C.H., Carlson,C. and Rohmann,G.F.  
 Identification, sequence, and transcriptional analysis of lef-3, a  
 gene essential for Orygia pseudotsugata baculovirus DNA replication  
 Virology 210 (2), 372-382 (1995)  
 95343549  
 MEDLINE  
 REFERENCE  
 3 (bases 1 to 5388)  
 Ahrens,C.H. and Rohmann,G.F.  
 The DNA polymerase and helicase genes of a baculovirus of Orygia  
 pseudotsugata



CDS  
 RAYLRNROAVLDLKLAEDYADTAAYMAQDOPEASSRHAFATLNRRLLLGVDDPAR  
 RNLNSVARIETLLRVYNDANAEVNVISGDIYEERXSKISISVQPEAQTPRASAQOQ  
 TSLPRQTSLEPQTOASLPQOTPEDQPEWSPSPVHTTPTALPQOTPPATDPFSR  
 TSEFVYVQKERAVIDTRKPPVPRPKHLKSRSSVATNAAGATPAVAPPPPSAD  
 VTTSMPPPPPSADVTTSMPPPPVADLTSMPPPPPVADLTSMPPPVADLTSMP  
 INNLLIDMAVAVETKNAGDNRSALLDOIKOGTKTKTOPADGAPATDPRSTLLSEIQ  
 GKTTLKRIKIDOSTOPLIKDVTDTDKTKLTKNFVNIIDRISKOBEEKDRIDTIT  
 KRPAVEHTDNGSTGNSDDMRD"  
 complement(2533..3270)  
 /note="ORF3: ph, AcMNPV ORF8 homolog"  
 /codon\_start=1  
 /product="polynhedrin"  
 /protein\_id="AAC59002.1"  
 /db\_xref="GI:1911249"  
 /translation="MPTSYRPTTGRITTYVYNYKTKKNGSVITKNAKKHLLHEEDE  
 KHLDPDHYMAVEDPFLGPKKQKLTLEKRNKVPDTMKLIVMNSGEELREITR  
 VDSFPIYDQEMDVFLVNNRPTPNPKYFQAOLHMLDCCDQVPEHVRIVEPSY  
 VGMNNEYRISLAKGGGCPINNIAEYTNSEFSFVNRYIWEYKPIYVIGTDSSE  
 ELLIEYSLVFKYKEFAPADPLFTGPAY"  
 complement(3350..3541)  
 /note="ORF4"  
 /codon\_start=1  
 /product="unknown"  
 /protein\_id="AAC59003.1"  
 /db\_xref="GI:1911250"  
 /translation="MRKNNGASQYPEHLTTVAKRHRRLQSLCVCWCDQKSWDPHSAKGL  
 MEHMFCCURITYOQKIKH"  
 complement(3698..4093)  
 /note="ORF5"  
 /codon\_start=1  
 /product="unknown"  
 /protein\_id="AAC59004.1"  
 /db\_xref="GI:1911251"  
 /translation="MGASQNSECKITKLEFYRNSQGAALDEKDLHCLYDLREYGA  
 HLNRADDAARRKCKAEKALKRVEITADPRHMLEAATAACADGCMSTLSAQDDI  
 AREKEIVDRITRQLKQDSLKQDRLLKR"  
 complement(4125..4739)  
 /note="ORF6: left-2, AcMNPV ORF6 homolog"  
 /codon\_start=1  
 /product="late expression factor 2"  
 /protein\_id="AAC59005.1"  
 /db\_xref="GI:1911252"  
 /translation="MERVWNPAGIDGLKSEITYLVDPHDFGVGLTSLPYFERGLF  
 VRMSGRLIALAARPOEPQARRPQSRNRNVCIKACADGASQSLAVLARVSM  
 PMSKTMADISSAPRGMYRKREPCNYLANVITCTKACIALGALLHFTYMDACV  
 GEVTHLIKADQYVYKSNCKMKKVTYKLCQASMCGLNPICNF"  
 complement(4742..4975)  
 /note="ORF7: AcMNPV ORF5 homolog"  
 /codon\_start=1  
 /product="unknown"  
 /protein\_id="AAC59006.1"  
 /db\_xref="GI:1911253"  
 /translation="MNRPTVNRNTAAVYTDYDRQRLRELNSLRSHHELCTRSATGFD  
 CNRPLEAGDRAPAVYAAANGGSHSLIDCKV"  
 5020..5460  
 /note="ORF8: AcMNPV ORF4 homolog"  
 /codon\_start=1  
 /product="unknown"  
 /protein\_id="AAC59007.1"  
 /db\_xref="GI:1911254"  
 /translation="MGAARFVRSLSRTIOFERKENVAVHVDHMLGRLALIDKTSADY  
 RRRGFLSRNVLSACMAVNVQVAPDATIMRQPTIYFVNCQCHAMADVPAPDDHS  
 IARYLAECAVAVLIDHPLDVFGETEGVNELLEVORINAGDGL"  
 complement(5669..6151)  
 /note="ORF9: ptp-2"  
 /codon\_start=1  
 /product="protein tyrosine phosphatase 2"  
 /protein\_id="AAC59008.1"  
 /db\_xref="GI:1911255"  
 /translation="MYEASQIDEHVFVGYGDNEMALRIENHAIASVSLSDSVA  
 PIRAHGLPVEDHITHVCEAPRTCAALPNNAPALDYVRRIRGEGRVLAHCYAGSR  
 SNAALAVYILMSROMPYQDALNAVQSKRAVAINHVFRLATRCSTYFVNDLKPQVY"

CDS  
 complement(6129..6791)  
 /note="ORF10: ptp-1, AcMNPV ORF1 homolog"  
 /codon\_start=1  
 /product="protein tyrosine phosphatase 1"  
 /protein\_id="AAC59009.1"  
 /db\_xref="GI:1911256"  
 /translation="MLTSRMAVITYRTSERNGGACWTCNGVRRRQVHLPSVRAM  
 FPRMHEYTAGCAVLEGTRLCEKYLEEYVTSDEDRWTASVLAHRBALAVI  
 DLNTFARYDGAQMVKMLYKIRVGRVAPDDDLVAFIEITVEEFPKCPMTVAV  
 HHTHGLNRSGLVCRVMEVERLGVSTDAIARETAGHKIERTNYLQDILLAKRHVQ  
 PN"  
 6819..7814  
 /note="ORF11: AcMNPV ORF11 homolog"  
 /codon\_start=1  
 /product="unknown"  
 /protein\_id="AAC59010.1"  
 /db\_xref="GI:1911257"  
 /translation="MSLSKLLVYAYGYNLPDHRGSGYHLXRTVHEHLNTYVSN  
 ASCVRDIATVRCNLNSGHDFDVAROLDVSEVAAITLSMFRGCGATGCGAMQVLA  
 DIDHAPLARVGRKRANIETALDIADIPSDVNTNLTIGREKHIFRSCGLRVADV  
 DPIDRADGMMYHRECVLTYMHLVACGAVPAGSATRLDAAKHIQFNDGNCAPATAV  
 VYGRFCAIGREHFAHKTACMHILFQMRNDLTPADERHPCFVIGIDFRQCKDYTD  
 LRPHADALYIHGTTDRQNALFDLCCVNASDIDADCYDCVVKFYATONKKYKM"  
 complement(8180..9142)  
 /note="ORF12: AcMNPV ORF13 homolog"  
 /codon\_start=1  
 /product="unknown"  
 /protein\_id="AAC59011.1"  
 /db\_xref="GI:1911258"  
 /translation="MFSMFGMNNNADEOVNAERDEQAVRYAVDORAHSDLVRRDVF  
 KCHPFVFRERVYDETRGCCSYVDPRCKGIKSHIDLSPROHVRLOLNEVLGAP  
 PAQDPSLGSLEFATKHLGVLQLOQSFANKREYLLAVANGMDHNDRLNLTAVLNHV  
 KTLNTNSDKFISAKRSKLEVGAEFOEDRLQTLKALDCAPTRAPVCPVPR  
 DVTGHPHLAVFMGRVEDRGVQIAPFAGQEHKPKRLEEESGDVDVRRANPLLA  
 VHCJKEEPANGCHKIRRLPKRPIEVDCAVNAVAKIYKALINT"  
 complement(9094..9825)  
 /note="ORF13: left-1, AcMNPV ORF14 homolog"  
 /codon\_start=1  
 /product="late expression factor 1"  
 /protein\_id="AAC59012.1"  
 /db\_xref="GI:1911259"  
 /translation="MAPCKYPERVQMMMDAIAVNSRRLAFMTDRPMVHAHNEFDS  
 AAOLPAYTVKNSISDVHYKPLEEGREVDADKDCDAKVAELKYNVGAFTMELF  
 ESKEDAVORINPSSNRGPHMLKFCGKFKMDAPSLSEHMPNVYKOPAKLYSGIRG  
 SFADVVRNAVMTIGDAREDDVLYKRPVDRDVCNNKQIRAPFSYNYGQY3RCL  
 TQQLQORIKGSAGCLAGTPTSK"  
 9893..11362  
 /note="ORF14: egt, AcMNPV ORF15 homolog"  
 /codon\_start=1  
 /product="ectyosteroid UDP-glucosyltransferase"  
 /protein\_id="AAC59013.1"  
 /db\_xref="GI:1911260"  
 /translation="MVFLLIALTLTAGARAASILAVLPYPAISHHYVYRAYALYK  
 NCHNTYVIRKPLADYAVODEGRVQIDADMSAQYKDLVASSGVFERKRGVADETTV  
 TADNYMGLILEMKDOEDNANVRRFSTRTDVAVVEAFADYALVFGHLEPPAVIQI  
 APGVGLAENFERERRVARPHLHYPTFGAALTRRGALSEWRLNEFELARRSDELL  
 KOQFKSTPTITQLEDNYOILLNHPYVDNAPRPSVOYLGGGHLAOLPORDLA  
 ELERLINESVDGAVYVSFGSGIDTNSHTAFELQMLDPEAFNLNNTYMKWDAAVVS  
 VALPRNVLAQKWFQTSOTAVLRKNVAVFTQAGLGSDDALDARPAWNCITPMKGOEFH  
 AKLQOFGVARALDITAAVASAPOLQLAIEVLAIDGASVYARIDKLRAVVEHDAADDERA  
 VKFERVYIKFNNDVWMPARSLSKTTAANNAYSIDYFVRPL"  
 11547..12140  
 /note="ORF15: AcMNPV ORF16 homolog"  
 /codon\_start=1  
 /product="unknown"  
 /protein\_id="AAC59014.1"  
 /db\_xref="GI:1911261"  
 /translation="METADPPISTVAPPKGAVCAVRTVTTTYSDSGNNEDRLIQ"

Query Match 72.48; Score 21; DB 14; Length 131995;  
 Best Local Similarity 82.88; Pred. No. 2e+02;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

Oy      1  caccgaagcgcgcgcgcgcgcgcac 29
         | | | | | | | | | | | | | | | | | |
Db      55017  cccgcaaaaggcgccgcccacgacgac 54989

RESULT  6
SC7H1/c 35654 bp  DNA  linear  BCF 22-MAR-2000
LOCUS Streptomyces coelicolor cosmid 7H1.
DEFINITION A1021411
VERSION A1021411.1 GI:2808758
KEYWORDS abfB; aminotransferase; arabinofuranosidase; cbfM; cbfN; cbfO;
          cbfQ; chitinase; cobalt transport system; export protein;
          oxidoreductase; peptidylprolyl isomerase; sigma factor;
          transferase; xlnA; xylanase A.
SOURCE Streptomyces coelicolor A3(2).
        Streptomyces coelicolor A3(2).
        Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
        Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 35654)
AUTHORS Redenbach,M., Kleser,H.M., Denapate,D., Eichner,A., Cullum,J.,
        Kinasli,H. and Hopwood,D.A.
        A set of ordered cosmids and a detailed genetic and physical map
        for the 8 Mb Streptomyces coelicolor A3(2) chromosome
        Mol. Microbiol. 21 (1), 77-96 (1996)
        97000351
        2 (bases 1 to 35654)
        Murphy,L. and Harris,D.
        Unpublished
        3 (bases 1 to 35654)
        Parkhill,J., Barrett,B.G. and Rajandream,M.A.
        Direct Submission
        Submitted (20-JAN-1998) Streptomyces coelicolor sequencing project,
        Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
        CV10 1SA E-mail: barrellesanger.ac.uk Cosmids supplied by Prof.
        David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
        Colney, Norwich, Norfolk NR4 7QH, UK
        Notes:
        Streptomyces coelicolor sequencing at The Sanger Centre is funded
        by the BBSRC.
        Details of S. coelicolor sequencing at the Sanger Centre are
        available on the World Wide Web.
        (URL: http://www.sanger.ac.uk/Projects/S.coelicolor/) CDS are
        numbered using the following system eg SC7B7.01c. SC (S.
        coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
        strand).
        The more significant matches with motifs in the PROSITE database
        are also included but some of these may be fortuitous. The length
        in codons is given for each CDS.
        Usually the highest scoring match found by fasta -o is given for
        CDS which show significant similarity to other CDS in the database.
        The position of possible ribosome binding site sequences are given
        where these have been used to deduce the initiation codon. Gene
        prediction is based on positional base preference in codons using a
        specially developed Hidden Markov Model (Krogh et al., Nucleic
        Acids Research, 22(22):4768-4778(1994)) and the Frameplot program
        of Bibb et al., Gene 30:157-66(1984) as implemented at
        http://www.nh.90.jp/
        jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
        correct initiation codon. Where possible we choose an initiation
        codon (atg, gtg, ttg or att) which is preceded by an upstream
        ribosome binding site sequence (optimally 5-13bp before the
        initiation codon). If this cannot be identified we choose the most
        upstream initiation codon.
        IMPORTANT: This sequence MAY NOT be the entire insert of the
        sequenced clone. It may be shorter because we only sequence
        overlapping sections once, or longer, because we arrange for a
        small overlap between neighbouring submissions. Cosmid 7H1 lies
        between 10A5 and 7B7 on the AseI-B genomic restriction fragment.
        Location/Qualifiers
            1..35654
            /organism="Streptomyces coelicolor A3(2)"
            /strain="A3(2)"

```

```

misc_feature
gene
CDS

/db_xref="taxon:100226"
/clone="cosmid 7H1"
/complement(1..441)
/gene="xlnA"
/complement(1..189)
/gene="xlnA"
/note="Pfam match to entry glycosyl_hydro3 PF00331,
glycosyl hydrolases family 10, from 1 to 71, score 107.60"
/complement(1..441)
/gene="xlnA"
/note="SC7H1.01c, xlnA, xylanase A precursor, partial CDS,
len >147 aa; almost identical to S. lividans xlnA_STRLT
P26514 endo-1,4-beta-xylanase A precursor (ec 3.2.1.8)
(XYLASE A) (477 aa), fasta scores; opt: 926 z-score:
1591.1 E(): 0, 97.9% identity in 146 aa overlap. Contains
N-terminal signal sequence and Pfam match to entry
glycosyl_hydro3 PF00331, Glycosyl hydrolases family 10,
score 107.60"
/codon_start=1
/translation="MGSYALPRSCYRSIRVLALVGVLTATLALPPGAHAES
TLGAARQSGREFTALSGRLSDSTYSISLAGRENNVTAKENKIDATBPRQGFNF
SSADRYVMVAQNGKQVGRHTLAH50QPGMWSLGSALROAMI"
/complement(1..2585)
/note="96% match to EM_BA:SLXLA M6451 Streptomyces
lividans arabinofuranosidase precursor (abfB) and xylanase
A precursor (xlnA) genes from: 166 to: 2759"
1..1158
/note="overlap with cosmid St10A5"
/complement(449..453)
/note="possible RBS upstream of xlnA"
823..826
/note="possible RBS upstream of abfB"
832..2259
/gene="abfB"
832..2259
/note="SC7H1.02, abfB, arabinofuranosidase precursor, len:
475 aa; almost identical to S. lividans TR:P96463
(EMBL:M6451) arabinofuranosidase precursor (EC
3.2.1.55) (alpha-1-arabinofuranosidase) (arabinosidase)
(478 aa), fasta scores; opt: 3129 z-score: 1807.5 E(): 0,
95.8% identity in 477 aa overlap. Contains N-terminal
signal sequence"
/codon_start=1
/translation="MHRGSLSRGHTSAVLAAYVAALALALVATTPAQAAGSGLR
GAGSNRDLVLGSDGDGALLDLXDMGCTNQOQWTSMDGRLLVYDCKDIPGHAHA
PGRVRYWCSGCAANOQRNRSQTVYGVSGICLDAAGATNGTAIVDLCVGGGN
OKRTGLCTPTDGTALPSTYRWSSTGVIAQPSVVALKDTVTYHNGRHLYVGSF
SSGSSYSWSEPTTNWSDASAGQANMAVAAPLTFAPKNIVLAVYONGSWPFI
YRNSDPTDNGWSADPOLPTFGSISGSDGPIDQTLADGONMYLFPADNGKIVAS
MPTGNFPTNGSSYTTIMSDTKANLFEQVGVYVVOCONYIMTVEAMGANGRFRFT
ASISGSMWTOQAASEGNPFAGKANSGLATWINDSHDLDVRDNDQTMVDPCLQPLY
OGKSPNAGGYNLSLPMWPGVLTLR"
/complement(2243..3418)
/gene="SC7H1.03c"
/complement(2243..3418)
/note="SC7H1.03c, unknown, possible membrane protein, len:
391 aa; contains PS00213 Lipocalin signature, though this
may be spurious"
/codon_start=1
/translation="
/product="hypothetical protein SC7H1.03c"

```





Estimated insert size: 122355; sum-of-contigs estimation  
 Quality coverage: 0x in Q20 bases; agarose-fp estimation  
 Quality coverage: 2.1x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 NOTE: This is a 'working draft' sequence. It currently  
 consists of 74 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

|   |       |                                    |
|---|-------|------------------------------------|
| 1 | 5209: | contig of 5209 bp in length        |
| * | 5210  | 5309: gap of unknown length        |
| * | 5310  | 10338: contig of 5029 bp in length |
| * | 10339 | 10438: gap of unknown length       |
| * | 10439 | 14826: contig of 4388 bp in length |
| * | 14827 | 14926: gap of unknown length       |
| * | 14927 | 19138: contig of 4212 bp in length |
| * | 19139 | 19238: gap of unknown length       |
| * | 19239 | 23107: contig of 3869 bp in length |
| * | 23108 | 23207: gap of unknown length       |
| * | 23208 | 26969: contig of 3762 bp in length |
| * | 26970 | 27069: gap of unknown length       |
| * | 27070 | 32006: contig of 4937 bp in length |
| * | 32007 | 32106: gap of unknown length       |
| * | 32107 | 34680: contig of 2574 bp in length |
| * | 34681 | 34780: gap of unknown length       |
| * | 34781 | 37409: contig of 2629 bp in length |
| * | 37410 | 37509: gap of unknown length       |
| * | 37510 | 41235: contig of 3726 bp in length |
| * | 41236 | 41335: gap of unknown length       |
| * | 41336 | 44424: contig of 3089 bp in length |
| * | 44425 | 44524: gap of unknown length       |
| * | 44525 | 48770: contig of 4246 bp in length |
| * | 48771 | 48870: gap of unknown length       |
| * | 48871 | 52031: contig of 3161 bp in length |
| * | 52032 | 52131: gap of unknown length       |
| * | 52133 | 55038: contig of 2907 bp in length |
| * | 55039 | 55138: gap of unknown length       |
| * | 55139 | 58265: contig of 3127 bp in length |
| * | 58266 | 58365: gap of unknown length       |
| * | 58366 | 61071: contig of 2706 bp in length |
| * | 61072 | 61171: gap of unknown length       |
| * | 61172 | 62802: contig of 1631 bp in length |
| * | 62803 | 62902: gap of unknown length       |
| * | 62903 | 65216: contig of 2314 bp in length |
| * | 65217 | 65316: gap of unknown length       |
| * | 65317 | 68673: contig of 3357 bp in length |
| * | 68674 | 68773: gap of unknown length       |
| * | 68774 | 71777: contig of 3004 bp in length |
| * | 71778 | 71877: gap of unknown length       |
| * | 71878 | 75040: contig of 3163 bp in length |
| * | 75041 | 75140: gap of unknown length       |
| * | 75141 | 77111: contig of 1971 bp in length |
| * | 77112 | 77211: gap of unknown length       |
| * | 77212 | 78927: contig of 1716 bp in length |
| * | 78928 | 79027: gap of unknown length       |
| * | 79028 | 81095: contig of 2068 bp in length |
| * | 81096 | 81195: gap of unknown length       |
| * | 81196 | 83729: contig of 2534 bp in length |
| * | 83730 | 83829: gap of unknown length       |
| * | 83830 | 85747: contig of 1918 bp in length |
| * | 85748 | 85847: gap of unknown length       |
| * | 85848 | 88908: contig of 3061 bp in length |
| * | 88909 | 89008: gap of unknown length       |
| * | 89009 | 90558: contig of 1550 bp in length |
| * | 90559 | 90658: gap of unknown length       |
| * | 90659 | 93270: contig of 2612 bp in length |
| * | 93271 | 93370: gap of unknown length       |
| * | 93371 | 95096: contig of 1726 bp in length |

|       |        |        |                                     |
|-------|--------|--------|-------------------------------------|
| 95097 | 95197  | 96411  | 96411: gap of unknown length        |
| *     | 96412  | 96741  | 96741: gap of unknown length        |
| *     | 96742  | 98704  | 98704: contig of 1963 bp in length  |
| *     | 98705  | 98804  | 98804: gap of unknown length        |
| *     | 98805  | 100546 | 100546: contig of 1742 bp in length |
| *     | 100547 | 100646 | 100646: gap of unknown length       |
| *     | 100647 | 101882 | 101882: contig of 1236 bp in length |
| *     | 101883 | 101982 | 101982: gap of unknown length       |
| *     | 101983 | 103008 | 103008: contig of 1026 bp in length |
| *     | 103009 | 103108 | 103108: gap of unknown length       |
| *     | 103109 | 104820 | 104820: contig of 1712 bp in length |
| *     | 104821 | 104921 | 104921: gap of unknown length       |
| *     | 104922 | 106395 | 106395: contig of 1475 bp in length |
| *     | 106396 | 106496 | 106496: gap of unknown length       |
| *     | 106497 | 108468 | 108468: contig of 1973 bp in length |
| *     | 108469 | 108568 | 108568: gap of unknown length       |
| *     | 108569 | 110201 | 110201: contig of 1633 bp in length |
| *     | 110202 | 110301 | 110301: gap of unknown length       |
| *     | 110302 | 111838 | 111838: contig of 1537 bp in length |
| *     | 111839 | 111938 | 111938: gap of unknown length       |
| *     | 111939 | 113656 | 113656: contig of 1718 bp in length |
| *     | 113657 | 113756 | 113756: gap of unknown length       |
| *     | 113757 | 115783 | 115783: contig of 2027 bp in length |
| *     | 115784 | 115883 | 115883: gap of unknown length       |
| *     | 115884 | 117732 | 117732: contig of 1849 bp in length |
| *     | 117733 | 117832 | 117832: gap of unknown length       |
| *     | 117833 | 119791 | 119791: contig of 1959 bp in length |
| *     | 119792 | 119891 | 119891: gap of unknown length       |
| *     | 119892 | 121704 | 121704: contig of 1813 bp in length |
| *     | 121705 | 121804 | 121804: gap of unknown length       |
| *     | 121805 | 123619 | 123619: contig of 1815 bp in length |
| *     | 123620 | 123719 | 123719: gap of unknown length       |
| *     | 123720 | 125324 | 125324: contig of 1605 bp in length |
| *     | 125325 | 125424 | 125424: gap of unknown length       |
| *     | 125425 | 127220 | 127220: contig of 1796 bp in length |
| *     | 127221 | 127320 | 127320: gap of unknown length       |
| *     | 127321 | 128937 | 128937: contig of 1617 bp in length |
| *     | 128938 | 129037 | 129037: gap of unknown length       |
| *     | 129038 | 130061 | 130061: contig of 1024 bp in length |
| *     | 130062 | 130161 | 130161: gap of unknown length       |
| *     | 130162 | 131556 | 131556: contig of 1395 bp in length |
| *     | 131557 | 131656 | 131656: gap of unknown length       |
| *     | 131657 | 132821 | 132821: contig of 1165 bp in length |
| *     | 132822 | 132921 | 132921: gap of unknown length       |
| *     | 132922 | 134359 | 134359: contig of 1438 bp in length |
| *     | 134360 | 134459 | 134459: gap of unknown length       |
| *     | 134460 | 136156 | 136156: contig of 1697 bp in length |
| *     | 136157 | 136256 | 136256: gap of unknown length       |
| *     | 136257 | 137579 | 137579: contig of 1323 bp in length |
| *     | 137580 | 137680 | 137680: gap of unknown length       |
| *     | 137681 | 138787 | 138787: contig of 1108 bp in length |
| *     | 138788 | 138887 | 138887: gap of unknown length       |
| *     | 138888 | 140046 | 140046: contig of 1159 bp in length |

Query Match 69.0%; Score 20; DB 2; Length 162623;  
 Best Local Similarity 82.1%; Pred. No. 4; Se+02;  
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

07 2 acgcgaagcgctcgccagcaagac 29  
 db 110089 ACGCCGGGGCGTCGCGCACCCACGAC 110062

RESULT 8 BC011138 2053 bp mRNA linear PRI 30-JUL-2001  
 LOCUS BC011138/c  
 DEFINITION Homo sapiens, clone MGC:17456 IMAGE:3448753, mRNA, complete cds.  
 ACCESSION BC011138  
 VERSION BC011138.1 GI:15029829  
 KEYWORDS MGC.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 2053)  
Strausberg, R.  
Direct Submission  
Submitted (25-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550, USA

REMARK  
COMMENT

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ARCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNI)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [villalobcm.tmc.edu](mailto:villalobcm.tmc.edu)  
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNI at: <http://image.llnl.gov>  
Series: IRAC Plate: 19 Row: 9 Column: 13.  
Location/Qualifiers

FEATURES  
source

1. 2053  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:17456 IMAGE:3448753"  
/tissue\_type="Placenta, choriocarcinoma"  
/clone\_lib="NIH\_MGC\_10"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
459. 1079  
/codon\_start=1  
/product="Unknown (protein for MGC:17456)"  
/protein\_id="AAH1118.1"  
/db\_xref="GI:15029830"  
/translation="MTPEHPLTTSRRPAPQPCARCAPAVEKDPSPSPARTHTPR  
EPLHAGRGSPSSGTGPRGSGHSGQAHVSGSLAQPCALGSDGLSSQPAH  
LAPGTHSPAPLRLRTSPRHPRGSGSGVSTMMASRDGITYLSAOPPEATP  
VLEGSRAVALPAALPERRACCTSSRCGPRARPRGQPPSAE"

## CDS

BASE COUNT  
ORIGIN

404 a 604 c 559 g 406 t

Query Match 68.3%; Score 19.8; DB 9; Length 2053;  
Best Local Similarity 91.3%; Pred. No. 1.2e+03;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 gcaagagcgctgcgccaagcacc 26  
|||||  
Db 1114 GCAAGAGCGCGTCGCGGCGCAC 1092

RESULT 9  
AF357175/c 4926 bp DNA linear BCT 09-MAY-2001  
LOCUS  
DEFINITION Mycobacterium tuberculosis isolate DS13 unknown sequence.  
ACCESSION AF357175  
VERSION AF357175.1 GI:14009932  
KEYWORDS  
SOURCE Mycobacterium tuberculosis.  
ORGANISM Mycobacterium tuberculosis  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Mycobacteriaceae;  
Mycobacterium; Mycobacterium tuberculosis complex.  
REFERENCE 1 (bases 1 to 4926)  
Small, P.M.  
Comparing genomes within the species Mycobacterium tuberculosis

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Unpublished  
2 (bases 1 to 4926)  
Small, P.M., Kato-Maeda, M., Rhee, J.T., Gingeras, T.R., Salamon, B., Drenkow, J., and Smittipat, N.  
Direct Submission  
Submitted (07-MAR-2001) Medicine, Division of Infectious Diseases and Geographic Medicine, Stanford University, 300 Pasteur Dr., Grant Building S 143, Stanford, CA 94305, USA

FEATURES  
source

1. 4926  
/organism="Mycobacterium tuberculosis"  
/strain="H37Rv"  
/isolate="DS13"  
/db\_xref="taxon:1773"  
1. 4926  
/note="sequence found to be deleted in some clinical isolates"

BASE COUNT  
ORIGIN

992 a 1473 c 1552 g 909 t

Query Match 68.3%; Score 19.8; DB 1; Length 4926;  
Best Local Similarity 91.3%; Pred. No. 1e+03;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 cgaagagcgctgcgccaagcacc 25  
|||||  
Db 932 CGCAAGCGCGCTCGGCCATCCA 910

RESULT 10  
AB011161/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AB011161 5047 bp mRNA linear PRI 10-APR-1998  
Homo sapiens mRNA for KIAA0589 protein, partial cde.  
AB011161.1 GI:3043701  
KIAA0589 protein.  
Homo sapiens male brain cDNA to mRNA, clone\_lib:pbjuescriptII SK plus clone:HJ2695.  
Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 5047)  
Ohara, O., Nagase, T., and Ishikawa, K.  
Direct Submission  
Submitted (13-FEB-1998) Osamu Ohara, Kazusa DNA Research Institute, DNA Technology; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail: [cdnainfo@kazusa.or.jp](mailto:cdnainfo@kazusa.or.jp), Tel: +81-438-52-3913, Fax: +81-438-52-3914)  
2 (sites)  
Nagase, T., Ishikawa, K., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O.  
Prediction of the coding sequences of unidentified human genes. IX. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro  
DNA Res. 5 (1), 31-39 (1998)  
98290545

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
FEATURES  
source

1. 5047  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HJ2695"  
/sex="male"  
/tissue\_type="brain"  
/clone\_lib="pbjuescriptII SK plus"  
1. 2064  
/gene="KIAA0589"  
/l. 2064  
/gene="KIAA0589"  
/codon\_start=1  
/product="KIAA0589 protein"  
/db\_xref="GI:3043702"

## CDS

gene  
1. 2064  
/gene="KIAA0589"  
/l. 2064  
/gene="KIAA0589"  
/codon\_start=1  
/product="KIAA0589 protein"  
/db\_xref="GI:3043702"





repeat\_region /rpt\_family="AluSc"  
745..911  
/rpt\_family="Alu"  
repeat\_region 1803..1993  
/rpt\_family="MER20"  
complement(join(<2487..2603,2694..2856,3741..3886,  
4321..4416,7302..>7579))  
/note="Hypothetical partial protein; Most similar to  
YB3C-SCHPO HYPOTHETICAL 60.9 KD PROTEIN C2F12.12C IN  
CHROMOSOME II gnl|PIID|e325370 (297211) hypothetical  
protein [Schizosaccharomyces pombe]; 5'-end of  
hypothetical gene not discernible in genomic sequence at  
this time"  
/codon\_start=3  
/evidence=not-experimental  
/product="R31449\_3"  
/protein\_id="AAC32903.1"  
/db\_xref="GI:3461802"  
/translation="RSGSPRAAAALSSQOSLQRLRLREKRKQOEELMKAFETPEEK  
BARLAKREAKRKREKMGVEYMGVTNDNPFQNNLGTETWNALEKGIISL  
EEKELERKRIQEDNRLQKQLRLERERKARQOELEMLQREKAEKFTWE  
QEDNFHQAKIRSKIRIDGRAPKIDILAKYISAEDDLAEHMEHPTFLNGLTVA  
MEDLEDIQVMELEQGNADFWMDMTTFDELSKLKLEASGKPG"  
misc\_feature complement(2489..2603)  
/note="Predicted exon, program: graal2exons\_human\_1.3,  
frame: 2, quality: excellent, score: 100.000"  
misc\_feature complement(2694..2856)  
/note="Predicted exon, program: graal2exons\_human\_1.3,  
frame: 2, quality: excellent, score: 100.000"  
misc\_feature complement(3741..3886)  
/note="Predicted exon, program: graal2exons\_human\_1.3,  
frame: 0, quality: excellent, score: 100.000"  
misc\_feature complement(4321..4416)  
/note="Predicted exon, program: graal2exons\_human\_1.3,  
frame: 1, quality: excellent, score: 99.000"  
repeat\_region complement(5352..5489)  
/rpt\_family="LINE2"  
5698..5997  
/rpt\_family="AluSx"  
repeat\_region 6002..6360  
/rpt\_family="MLTIC"  
repeat\_region 6568..6858  
/rpt\_family="AluDb"  
repeat\_region 6862..7152  
/rpt\_family="AluY"  
complement(7302..7579)  
/note="Predicted exon, program: graal2exons\_human\_1.3,  
frame: 2, quality: excellent, score: 96.000"  
misc\_feature complement(7478..7538)  
/rpt\_family="(CAG)n"  
repeat\_region complement(7819..8275)  
/rpt\_family="LINE2"  
repeat\_region complement(8516..8805)  
/rpt\_family="AluIo"  
repeat\_region 9062..9360  
/rpt\_family="AluSx"  
repeat\_region 9611..9994  
/rpt\_family="LINE2"  
repeat\_region complement(10187..10279)  
/rpt\_family="(CGG)n"  
repeat\_region complement(11027..11073)  
/rpt\_family="MIR"  
repeat\_region 11376..11675  
/rpt\_family="AluSx"  
repeat\_region 11697..11989  
/rpt\_family="AluSg"  
repeat\_region 12361..12559  
/rpt\_family="AluSx"  
repeat\_region complement(12660..12847)  
/rpt\_family="AluSg/x"  
repeat\_region 13102..13401  
/rpt\_family="AluSx"  
repeat\_region 13406..13703

misc\_feature /rpt\_family="AluY"  
13795..15474  
/note="DSD similarity to overlapping ESTs:  
(13795..14230) AA621023 ag3c09.s1 Soares testis NHT Homo  
sapiens cDNA clone 1056208 3'; Score: 856 Identity:  
432/436 (99%).-(13795..14267) A1018772 ov32e07.x1  
Soares testis NHT Homo sapiens cDNA clone IMAGE:1639044 3'  
similar to contains TAR1.t3 P785 repetitive element ;  
Score: 938 Identity: 471/473 (99%).-(14560..14100)  
AA128938 z011b12.r1 StrataGene neuroepithelium NT2RAMT  
937234 Homo sapiens cDNA clone 567359 5'; Score: 886  
Identity: 453/459 (98%).-(14913..14564) W94459 zei2e03.r1  
Soares fetal heart NBH19W Homo sapiens cDNA clone 358780  
5'. Score: 668 Identity: 345/350 (98%).-(15439..14978)  
R35135 y960e12.r1 Homo sapiens cDNA clone 36877 5'. Score:  
766 Identity: 447/476 (93%).-(15474..15189) AA323207  
EST25973 Cerebellum II Homo sapiens cDNA 5' end; Score:  
530 Identity: 278/284 (97%).-Additional EST matches:  
AA975069, AA128939, AA524185, AA483447, H67626, H73320,  
M78980, and many others."  
misc\_feature complement(13798..>37203)  
/note="Encodes human homolog of mouse  
phosphatidylinositol-4-phosphate 5-kinase I-gamma"  
misc\_feature complement(15645..16081)  
/note="DSD similarity to overlapping ESTs:  
(15842..15645) 242707.HSC06G021.H sapiens partial cDNA  
sequence; clone c-0rg02. Score: 388 Identity: 196/198  
(98%).-(15884..15653) T31346 EST34513 Homo sapiens cDNA 5'  
end similar to None; Score: 448 Identity: 228/232  
(98%).-(16081..15693) AA446775 zw89d06.r1 Soares total  
fetus NB2Hr8 9W Homo sapiens cDNA clone 784139 5'; Score:  
720 Identity: 377/387 (97%)."  
misc\_feature 16162..16753  
/note="DSD similarity to AA527155 n107c01.s1 NCI-CGAP\_Br2  
Homo sapiens cDNA clone IMAGE:967296; Score: 958  
Identity: 546/592 (92%)."  
CDS complement(join(16781..16783,17051..17134,22498..22630,  
25319..25423,26521..26553,26857..26995,27701..27865,  
29588..29672,30952..31000,32239..32322,35440..35645,  
36904..>37203))  
/function="putative lipid kinase"  
/note="Human homolog of mouse phosphatidylinositol-  
4-phosphate 5-kinase I-gamma"  
misc\_feature /codon\_start=1  
/product="K1AA0589, partial CDS"  
/protein\_id="AAC32904.1"  
/db\_xref="GI:3461803"  
/translation="NLNQNPRLLPKFEGLYCYVSGGKNIKVYVNNNLLPVRVKNHLK  
FDLKGSTYKRRASKEKESPPYKQADPMDQDEGLLDPTFSALVKTORCLYL  
ESFKIMYSLLIGVNTDOHERROAGCAOSTSEKRPVCGKALYSTRAMESTOGGAAR  
GEATESDDTMGIPRAVNGRGRRLLLHGITIDILDSYRFLKLEHTWALVHDGTVSV  
HRPSVIAERPFKSNVFRKNSSLSKSPSKRGGAGALAKVGPAAASAQIPEE  
REEAQYDLRGARSYPTLEDEGRPDLCPTSPSEFATASLATTLSSTLSIPRPS  
ETSEQPRRRRTQSGGODRQPEEPEDQCTTVQVEPSCSEIVVPRKEDGVSVA  
SPAGASAAVEYETASQASDERGAPASQASDEBAPADVDIYPTPBERSMVYSPLHYSNO  
APASGDSGSPR"  
misc\_feature complement(18117..18182)  
/note="Predicted exon, program: graal2exons\_human\_1.3,  
frame: 2, quality: good, score: 52.000"  
misc\_feature complement(18733..18853)  
/note="Predicted exon, program: graal2exons\_human\_1.3,  
frame: 0, quality: good, score: 68.000"  
repeat\_region 18879..19014  
/rpt\_family="AluSx"  
repeat\_region 19016..19311  
/rpt\_family="AluSx"  
repeat\_region 19312..19613  
/rpt\_family="AluY"  
repeat\_region 19618..19790  
/rpt\_family="AluSx"  
repeat\_region 20316..20356

misc\_feature complement(21150..21227)

Query Match 68.3%; Score 19.8; DB 9; Length 38209;  
 Best Local Similarity 91.3%; Pred. No. 7.1e+02;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 gcaagcgctgcgcgcgcac 26  
 |||||||||||||||||  
 Db 14718 GCAAGCGCTGCgcgcgcac 14740

RESULT 13  
 AC068475 175625 bp DNA linear HTG 01-JUN-2000  
 LOCUS Homo sapiens chromosome 8 clone RP11-383B15 map 8, WORKING DRAFT  
 DEFINITION SEQUENCE, 38 unordered pieces.  
 AC068475  
 AC068475.2 GI:8140670  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 175625)  
 Birtten,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens chromosome 8, clone RP11-383B15  
 Unpublished  
 2 (bases 1 to 175625)  
 Birtten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
 Bonoslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,  
 Campiano,A., Castle,A., Chepel,Y., Colangelo,M., Collins,S.,  
 Collymore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
 Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,  
 McCarthy,M., McKean,P., McGuck,A., McKernan,K., Mcpheeters,R.,  
 Melidiri,J., Meneus,L., Milnova,T., Miranda,C., Mleaga,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neill,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Plesni,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Testaev,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,D.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jun 1, 2000 this sequence version replaced gi:7677958.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIRB  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: 19767  
 Center clone name: 383.B.15  
 ----- Summary Statistics  
 Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 15681 bases at least Q40  
 Consensus quality: 166139 bases at least Q30  
 Consensus quality: 169716 bases at least Q20  
 Insert size: 187000; agarose-1p

Insert size: 171925; sum-of-contings  
 Quality coverage: 3.7 in Q20 bases; agarose-1p  
 Quality coverage: 4.0 in Q20 bases; sum-of-contings

NOTE: This is a 'working draft' sequence. It currently  
 consists of 38 contings. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contings are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 1032: contig of 1032 bp in length  
 1033 1132: gap of 100 bp  
 1133 2147: contig of 1015 bp in length  
 2148 2247: gap of 100 bp  
 2248 3601: contig of 1354 bp in length  
 3602 3701: gap of 100 bp  
 3702 5049: contig of 1348 bp in length  
 5050 5149: gap of 100 bp  
 5150 6694: contig of 1545 bp in length  
 6695 6794: gap of 100 bp  
 6795 7668: contig of 1074 bp in length  
 7669 7968: gap of 100 bp  
 7969 9659: contig of 1691 bp in length  
 9660 9759: gap of 100 bp  
 9760 11604: contig of 1845 bp in length  
 11605 11704: gap of 100 bp  
 11705 13073: contig of 1369 bp in length  
 13074 13173: gap of 100 bp  
 13174 15155: contig of 1962 bp in length  
 15156 15255: gap of 100 bp  
 15256 17626: contig of 2371 bp in length  
 17627 17726: gap of 100 bp  
 17727 20315: contig of 2589 bp in length  
 20316 20415: gap of 100 bp  
 20416 21588: contig of 1173 bp in length  
 21589 21688: gap of 100 bp  
 21689 23898: contig of 2210 bp in length  
 23899 23998: gap of 100 bp  
 23999 26021: contig of 2023 bp in length  
 26022 26121: gap of 100 bp  
 26122 27520: contig of 1399 bp in length  
 27521 27620: gap of 100 bp  
 27621 29883: contig of 2263 bp in length  
 29884 29983: gap of 100 bp  
 29984 33054: contig of 3071 bp in length  
 33055 33154: gap of 100 bp  
 33155 36994: contig of 3840 bp in length  
 36995 37094: gap of 100 bp  
 37095 40456: contig of 3362 bp in length  
 40457 40556: gap of 100 bp  
 40557 44872: contig of 4316 bp in length  
 44873 44972: gap of 100 bp  
 44973 49433: contig of 4461 bp in length  
 49434 49533: gap of 100 bp  
 49534 54009: contig of 4476 bp in length  
 54010 54109: gap of 100 bp  
 54110 57886: contig of 3777 bp in length  
 57887 57986: gap of 100 bp  
 57987 63137: contig of 5151 bp in length  
 63138 63237: gap of 100 bp  
 63238 68261: contig of 5024 bp in length  
 68262 68361: gap of 100 bp  
 68362 73156: contig of 4795 bp in length  
 73157 73256: gap of 100 bp  
 73257 79722: contig of 6466 bp in length  
 79723 79822: gap of 100 bp  
 79823 86499: contig of 6677 bp in length  
 86500 86599: gap of 100 bp  
 86600 91151: contig of 4552 bp in length  
 91152 91251: gap of 100 bp  
 91252 98153: contig of 6902 bp in length

\* 98154 98253: gap of 100 bp  
\* 98254 103322: contig of 4969 bp in length  
\* 103223 103322: gap of 100 bp  
\* 103323 110113: contig of 6791 bp in length  
\* 110114 110213: gap of 100 bp  
\* 110214 116818: contig of 6605 bp in length  
\* 116819 116918: gap of 100 bp  
\* 116919 127763: contig of 10845 bp in length  
\* 127764 127863: gap of 100 bp  
\* 127864 137894: contig of 10031 bp in length  
\* 137895 137994: gap of 100 bp  
\* 137995 149790: contig of 11796 bp in length  
\* 149791 149890: gap of 100 bp  
\* 149891 175625: contig of 25735 bp in length.  
Location/Qualifiers  
1. 175625  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="8"  
/map="8"  
/clone="RP11-383b15"  
/clone\_11b="RPC1-11 Human Male BAC"  
1. 1032  
/note="assembly-fragment"  
1133. 2147  
/note="assembly-fragment"  
2248. 3601  
/note="assembly-fragment"  
3702. 5049  
/note="assembly-fragment"  
5150. 6694  
/note="assembly-fragment"  
6795. 7868  
/note="assembly-fragment"  
7969. 9659  
/note="assembly-fragment"  
9760. 11604  
/note="assembly-fragment"  
11705. 13073  
/note="assembly-fragment"  
13174. 15155  
/note="assembly-fragment"  
15256. 17626  
/note="assembly-fragment"  
17727. 20315  
/note="assembly-fragment"  
20416. 21588  
/note="assembly-fragment"  
clone\_end:T7  
vector\_side:right  
21689. 23898  
/note="assembly-fragment"  
23999. 26021  
/note="assembly-fragment"  
26122. 27520  
/note="assembly-fragment"  
clone\_end:SP6  
vector\_side:right  
27621. 29883  
/note="assembly-fragment"  
29884. 33054  
/note="assembly-fragment"  
33155. 36994  
/note="assembly-fragment"  
37095. 40456  
/note="assembly-fragment"  
40557. 44872  
/note="assembly-fragment"

Query Match 68.3%; Score 19.8; DB 2; Length 175625;  
Best Local Similarity 91.3%; Pred. No. 5.3e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 gcaaaagcgcgtcgcacacac 26  
|||||  
Db 118633 GCAAGGCGCGTCGCGCGGCAC 118655

RESULT 14  
AE004513/C  
LOCUS  
DEFINITION  
Pseudomonas aeruginosa PA01, section 74 of 529 of the complete  
genome.  
ACCESSION  
AE004513 AE004091  
VERSION  
AE004513.1 GI:9946669  
KEYWORDS  
SOURCE  
ORGANISM  
Pseudomonas aeruginosa.  
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
Pseudomonas.  
1 (bases 1 to 15470)  
Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,  
Hickey,M.J., Brinkman,F.S.L., Huftnagle,W.O., Kowalik,D.J., Lagrou,M.,  
Garber,R.L., Golltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,  
Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,  
Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.  
Complete genome sequence of Pseudomonas aeruginosa PA01, an  
opportunistic pathogen  
Nature 406 (6799), 959-964 (2000)  
20437337  
2 (bases 1 to 15470)  
Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,  
Hickey,M.J., Brinkman,F.S.L., Huftnagle,W.O., Kowalik,D.J.,  
Lagrou,M., Garber,R.L., Golltry,L., Tolentino,E.,  
Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,  
Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,  
Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Sater,M.H.,  
Hancock,R.E.W., Lory,S. and Olson,M.V.  
Direct Submission  
Submitted (16-MAY-2000) Department of Medicine and Genetics,  
University of Washington Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA

TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
20437337  
2 (bases 1 to 15470)  
Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,  
Hickey,M.J., Brinkman,F.S.L., Huftnagle,W.O., Kowalik,D.J.,  
Lagrou,M., Garber,R.L., Golltry,L., Tolentino,E.,  
Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,  
Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,  
Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Sater,M.H.,  
Hancock,R.E.W., Lory,S. and Olson,M.V.  
Direct Submission  
Submitted (16-MAY-2000) Department of Medicine and Genetics,  
University of Washington Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA

TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
20437337  
2 (bases 1 to 15470)  
Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,  
Hickey,M.J., Brinkman,F.S.L., Huftnagle,W.O., Kowalik,D.J.,  
Lagrou,M., Garber,R.L., Golltry,L., Tolentino,E.,  
Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,  
Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,  
Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Sater,M.H.,  
Hancock,R.E.W., Lory,S. and Olson,M.V.  
Direct Submission  
Submitted (16-MAY-2000) Department of Medicine and Genetics,  
University of Washington Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA

FEATURES  
Source  
1. 15470  
/organism="Pseudomonas aeruginosa"  
/strain="PA01"  
/db\_xref="taxon:287"  
126. 878  
/gene="PA0780"  
126. 878  
/gene="PA0780"  
/codon\_start=1  
/transl\_table=1  
/product="probable transcriptional regulator"  
/protein\_id="AAG04169.1"  
/db\_xref="GI:9946670"  
/translation="MFDTRLRLPDQSHTAHEHQLVMSLAGRAFEVNGCGEVC  
MRACIVPEAGHVFAGVDNRMILIDGENDPPEDDPLRLPEARYPLDADPN  
LISVAGAEIARVSGDPLTALGVLLRALHILRFEGCOPLAASGPIDLBIAHYR  
NIARITVAELAOYACLSPSHFHOEFDSCTPHQVLLINVRDCAARLLRESRLPV  
RIAEPCFSQSASATITMARXILGLTPSLKRG"  
complement(999. 3062)  
/gene="PA0781"  
complement(999. 3062)  
/gene="PA0781"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAG04170.1"  
/db\_xref="GI:9946671"  
/translation="MSSSGILFSPSPRLPLPLALACLIVSGETIGADGRSELP  
TANPLGNESPAFSSVLEDELTLROKSGETLNGIPGVSTYFGGASRPVIRGSD  
GDRILRLRNGVGLDASSLSYDAHVDPDPSVRLYRGPAALIVGGNIGVYNSF  
DNITPSRPVGHIGSGELRKGADTTSRSGALEPAGDNFALHVDASREFNVRIPG  
YAHSSRORIDSDGKHRVONSODGDCGAVGSGYHWEHGVAGLSYSGDVSNGSPAE  
DVLRLKMQDRIYAFASEIRDLSEGPFTSLKIDAAVTKYEHKEIEDGEGTTFKMEGYEG



```

gene
3526..6708
/ gene="puta"
/ note="PA0782"
3526..6708
/ gene="puta"
/ codon_start=1
/ transl_table=11
/ product="proline dehydrogenase Puta"
/ protein_id="AAG04171.1"
/ db_xref="GI:9946672"
/ translation="MFKASHVLOGEDQTRSAEFPPVISANYAVEAYIGELLQIAD
PGDGEIARRNALIETVRSRNDVDTLDLRLQYSLDTEGIMLCLAEALRYP
DAATADALIRKLSAADMQRHLGSDNVLNFAWGLVWTKVNIIDERTGRAPSVI
SLVQSGSEPYRAAMQAMKMKOPVIGRTISEALKNGRCRCOGTYTSPDLGEA
ALTRADEKRYMDYRQADITVGAEPYGPSPSISIKLSAHPRYEPAQREYLSL
FANYLEAVARKLVNGITIDAEADRLSELTEYKLRDPAIAGWGEELVYQAS
KROLPIVWMLTLIGRELGARIPRLVGAWDEIKOCOVAGDEPYPTREKDTDS
YLACARVLSERTRGVYPOFASHNATVTAIADAEARAGAGDEPFEORLHGNG
DALYDVIKRRNRVITYAPVGAHKDLPLTVRLLENGANSSFYHKLVDRPVETL
IQRPVQLQKFTLANDRIPLOAIFGAGKNSQGINNIONOMMELALAKPPLTQO
WQAPVYNGSKLSGPHHEKCPDLETVSVYASAEQAKALVLOAAMPKRSAP
VDERAGILERLADLEAQRGELMALCYLLEKSLQDIDEIREADVCRITYAQAQR
LYBEELRGPTGRNELFEGRGIFACVSPNPFPLAIFLQIISALVACNTVLAAPQ
TSLIARVLEMFEGALPKRYTALLPGDGLTGVCRDARVAGVAFSTGIDTARIQ
ROLAEKGAATLLETGONAMIVDSALPEOVIKDAVASFTSAGORCSALMVIY
OQDAERVTEILKGAAELKYPTREVSVDYPIVDAEAKIJEQHLAKAELKLA
ETKPYKGLNHHPYAPVAFISGIDELKKNPGVLLHVRYAODLEKVOAINTGG
LTMGVSHNEETARKIEELARVGNLYINRQIGAVGYCPFGHGSLSTGKACAGPY
LRFVSERTSVTFAVGNASLSLADAE
7020..8540
/ gene="putp"
/ note="PA0783"
7020..8540
/ gene="putp"
/ codon_start=1
/ transl_table=11
/ product="sodium/proline symporter Putp"
/ protein_id="AAG04172.1"
/ db_xref="GI:9946673"
/ translation="MSVNPPTLIPFVITYIAMVYLIGAAVRSNNPNDYILGSRSL
EVTALSGASDMGMLMGLPAVYLSLSLSWYALIGVAINMLPAGRIKVSRE
HNGNALILPYFTNRPEDSKRLTESALVILVFETTCASGIVAGARLPSTGSL
ETALMAGAAATIAVTFIIGFLAVSWTDIVQASLMIFLITIPYIWLATGVEPFTA
IELDATSPMLKGSFIVISLAWMGIGYFGPHILARMAADVSKSPARISMT
WMTICLGAVAGFPGIAYFOAHPEOAGVASENERYEIELAKILENPIAGLISAI
LAAVMTLSQILLVSSALTEDFVYAKFARKGASLELWVGAMVLLVAVIAWLSN
PENRVLGVSTAAVAGGAFGLVYFSLIKRMTRNALAGMVGAAVYIIMKNIIG
TGLYIIPGLFASVAIVFSLGKAPSTMLKRFDADEETREAHQSGAAPVYOR"
complement(8685..9614)
/ gene="PA0784"
complement(8685..9614)
/ gene="PA0784"
/ codon_start=1
/ transl_table=11
/ product="probable transcriptional regulator"
/ protein_id="AAG04173.1"
/ db_xref="GI:9946674"
/ translation="MRIPWASLPLTALRPFESAROSFRKAAAEAVTPRAVSHOI
RSLKAWLGVPLFORLPKARLIDCGERLFRSLKALLEISQVDTLRORSYGNLTVS
TTPAALWLVPLRGRIYAHPOINRLDTVCYVDLHQAASIDLVIRSLDDHNEY
GLCLDECFVAYGSPAVLAERQPTLISVWNSRLYAGWEAWCAOAGGDMWEGQ
PVIRQYDEEHAIQAAIAGGIVLASSILVSVASGLRPYADISVAGYSAICV
PGRHRHPVKAFFEMLRLEAROSGACRLERA"
9741..10379
/ gene="PA0785"
9741..10379

```

```

gene
10389..10694
/ gene="PA0786"
10389..10694
/ gene="PA0786"
/ codon_start=1
/ transl_table=11
/ product="probable acyl carrier protein phosphodiesterase"
/ protein_id="AAG04174.1"
/ db_xref="GI:9946675"
/ translation="MSRLIAVNASPRGERSQSRRLAEVLEAVLEAHQAQVAREYV
RVPLPAVTEAFVAAAFHPQPEQRSLAMQADALSDQVLELSDSLVSTPMNTSV
PSGLKANIDQIVRLGTFEDFVLDNGVAQYRPLLRGRKALIVSRGHPGFGNQAM
NHADPWLRTALFGIDETVVAABGERSGSRSPESCDEARQRLALARSA"
10389..10694
/ gene="PA0786"
10389..10694
/ gene="PA0786"
/ codon_start=1
/ transl_table=11
/ product="probable transporter"
/ protein_id="AAG04175.1"
/ db_xref="GI:9946676"
/ translation="MAVPHLVAAPAEVAFPMGKTSNGFGRIMPSLTVAAIGIY
FLTALRELPSVAVPIWTAISLQTVPLGLLGESTLAVKLVSGLVAGVGLK"
10750..11913
/ gene="PA0787"
10750..11913
/ gene="PA0787"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein"
/ protein_id="AAG04176.1"
/ db_xref="GI:9946677"
/ translation="MLTTVAASNYRTLDVMPRLNLTGANGAKSNVYRRLRL
ADSAAGEVPSLAEGSLSTLMAGEDISRAMKSGQAPLESSYRQSNVLELPGAGE
DGYCIDLGEPSQSKFGDLVIRECITMAGPLRANVLRSPPLRAGGGRN
EVAQHLATEPDSMGAADVPOVPELIGIRIRISNRPYHPSDAEAPARPGIGTH
PVLGADADLAALQYIREIGDAALDEADDAFPCSRILINPBGREVLIEQPL
LPRRTALSPGRTYRLITLALSPRPALLVINEPSTLHPLLALGRVIGVQAO
HSQVTVSHARILVATLEHRECHSLGKEDGEMRIQGLREIDRRAMVVR"
complement(11918..15043)
/ gene="PA0788"
complement(11918..15043)
/ gene="PA0788"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein"
/ protein_id="AAG04177.1"
/ db_xref="GI:9946678"
/ translation="MNCATGASMGASERTNDPSELQPCSLPASHPKSRRLQRL
MLIGAPLILAAVGIIVVHEAFSRLOARELRYATLDYEVROGSEALILPPADGPF
DRLRGYLLPSEMRLLDRDYAITROAHSPALMYAHRLEFPYAEKAQAGLIADG
KQVYIPDRYQRRYANPASPVLVQSLFTFNDRDLDERPVLNPAVDMGRFTQAA
LSQVGRMLGFSANSSGCGSTLATOIEKTRBSAGRTSGTKLRQMTSASVRSRQPA
NPARQODIVLVYLSVPLSAPGYEGVTLADGLMWGADYRQVGBALDGRAGLAQ
GLALRODILVIAHRSPVLAERGBDELDRMTDSHLVLTQAGVTEAPLRDALAQ
LAPRPSQPTVOPPLTNKVTLLARLGMVPLVDLRLDRLANTVLOHLEDSV
TAVLQKADPYAOLGILGERILPTSRVSVYFETLFEPTSGNVNRYQVNDTDP
PDINGSKLEIGSTAKLRVILASVLETVQAIHRDYGMSVAELRKVEVEPDLFLLMG
DIVASRDRDLISAMLOAMERKYSAPYSFPTGGLTFENFRKENDGSRMLLEAL
RSINLPEVRLRLDILRHDIYQNAKSVOLLADDDPPRAYLDRTYDKSQYTLERF
WYTRDKDANORLEFLDGLRPMVSVLAAIHRYLQPDQDLASSATLREKLVLEGSITD
KRAELYERYGKFNLDQGYARVHPLEMLGLYLOKQOATFGVAASGAERKE
VYGMLEKSRHKNARDKRILILEVEAFELDHQWKKLGYPPDHLVPSYATVAGSSGR

```

```

Query Match 67.6% Score 19.6; DB 1; Length 15470;
Best Local Similarity 84.6%; Pred. No. 1e+03;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 cagcgaaaggcgctggcgaaccac 26
||||| ||| ||||| |||
Db 3826 CAGCGAACAGCGCTGGCCAGACAC 3801

```

```

RESULT 15
AE008794

```



LOCUS AE008794 26591 bp DNA linear BCT 25-OCT-2001  
 DEFINITION *Salmonella typhimurium* LT2, section 102 of 224 of the complete genome.  
 ACCESSION AE008794 AE006468  
 VERSION AE008794.1 GI:16420649  
 KEYWORDS  
 SOURCE  
 ORGANISM  
*Salmonella typhimurium* LT2.  
*Salmonella typhimurium* LT2.  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; *Salmonella*.  
 REFERENCE  
 AUTHORS  
 1 (bases 1 to 26591)  
 Mclelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Portolick, S., All, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W., Stuckey, T., Nhan, M., Waterston, R. and Wilson, R.K.  
 Complete genome sequence of *Salmonella enterica* serovar *Typhimurium* LT2  
 JOURNAL  
 PUBMED  
 Nature 413 (6858), 852-856 (2001)  
 11677609  
 2 (bases 1 to 26591)  
 The *Salmonella typhimurium* Genome Sequencing Project.  
 AUTHORS  
 TITLE  
 Submitted (29-MAR-2001) Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA  
 JOURNAL  
 Supported by NIH grant 5U 01 A143283  
 COMMENT  
 Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs: GLIMMER: <http://www.tigr.org/softlab/glimmer/glimmer.html> and GeneMark: <http://opal.biology.gatech.edu/GeneMark/>  
 EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes: <http://www.genome.ad.jp/kegg/>, and Pedro Romero and Peter Karp at EcoCyc: <http://ecocyc.PangeaSystems.com/ecocyc/>  
 The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Saigado, Julio Collado-Vides and Reguondb; [http://kinch.cifn.unam.mx:8850/db/reguondb\\_intro.frameset](http://kinch.cifn.unam.mx:8850/db/reguondb_intro.frameset)  
 This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.  
 FEATURES  
 source  
 1. 26591  
 /organism="Salmonella typhimurium LT2"  
 /strain="LT2; SGSC 1412; ATCC 700720"  
 /db\_xref="ATCC:700720"  
 /db\_xref="taxon:99287"  
 /complement(68. .2238)  
 /gene="wzc"  
 /note="STM2118"  
 /complement(68. .2227)  
 /gene="wzc"  
 /note="Ortholog of E. coli orf, hypothetical protein (AAC75121.1); Blastp hit to AAC75121.1 (732 aa), 87% identity in aa 13 - 732"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative tyrosine-protein kinase in colanic acid export"  
 /protein\_id="AAL21020.1"  
 /db\_xref="GI:16420650"  
 /translation="MTEKAKSAAVTGSDEIDIGICVTEARWVUGTAIPALAA VITFEATPIYSADALVOIRONAGNSNLTINSALANKPPASDAEIOIISRVLVGR VDDLDIDIAVKNFPLPGAGWERLMGHNHNVKVFETPRETMSGOIFLTKVIGDKR YOLVSDGGFSAQGVAGPLNDKDVTRVEALDARPDTEFTVSKFSTGLMINNIONNLT

VTETGKDTGVLSTAFTEGDRDOIREILNSITRNLYOOD IARKSEACKSLAFLAKOLP  
 EYRSRLDYAEKNKNAFRODKSDVDPLEAKVAVLSMDVNDOLNEIFRKEAEISKLT  
 KAHPAVYRLLERKRALEDEKAKNGRTAMKPTQOEIVRLTRIDVESGOVYMOULNKO  
 QELKITEASTVGDRNIVDPATIPQGVLPKPKALITLGSITGLMLSTVGVLNLSLFR  
 GIESPOLAEHGIISVYATIPLESEMOKARDSVKTIKIKRYKOSOLIVAGNPTDAIEA  
 IRSKLTSLHFAMMOANNVLMITGVSPSIGFTFYCANLAAYISOTHRVLIIDCDMK  
 GYTHELLGTNNVDGLSDILAGKEIASCAPALANPDLIPRGVPPNSPRLSEEP  
 GELIMASSRDVLDIPPLIAYTDAIVERRHAGTLPIMARVAVNTLKEVEYSLSRF  
 DONGTGVGVILNSTFRATGCGDYGYEYEGSDSK"  
 complement(2224. .2685)  
 /gene="wzb"  
 /note="STM2117"  
 /complement(2224. .2673)  
 /gene="wzb"  
 /EC\_number="3.1.3.48"  
 /note="Ortholog of E. coli probable protein-tyrosine-phosphatase (AAC75122.1); Blastp hit to AAC75122.1 (147 aa), 83% identity in aa 1 - 146"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative protein-tyrosine-phosphatase in colanic acid export"  
 /protein\_id="AAL21021.1"  
 /db\_xref="GI:16420651"  
 /translation="MFNKILVYCVGNCVCRSPTAERLLKRFPSLTVASAGLAVGK  
 ADPAASVYSAHDLSENHCAROIARSRLCRDYDILMERHITAQDIDAPBRGKVM  
 LFGHWDSEIREIDPDYRKSQDAFEAVYLLERSAQNQAALNAEGKP"  
 complement(2233. .2238)  
 /gene="wzc"  
 /note="putative RBS for wzc; Reguondb: STMS1H002262"  
 /complement(2679. .3945)  
 /gene="wza"  
 /note="STM2118"  
 /complement(2679. .3818)  
 /gene="wza"  
 /note="Ortholog of E. coli putative polysaccharide export protein (AAC75123.1); Blastp hit to AAC75123.1 (379 aa), 93% identity in aa 1 - 379"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative polysaccharide export protein, outer membrane"  
 /protein\_id="AAL21022.1"  
 /db\_xref="GI:16420652"  
 /translation="MMSKMKLMPILASLISGCVLPSSNMSTGKXVYIKQODAF  
 DLDMMVAVYPLTEPLVQLRPRNVPAPNSIDQELASVYRGPEDVAVTYWDPRE  
 LTPPAGYRRESSDGMNMGODGTFYPTIKVSYVGTITSEISDITGRVAKTIADPQ  
 GVKNVIAFRSQKATISGVNKSQOATITNPLTVLDATVAGGLTMMAMRVNVLTHN  
 GKQRISLQALMONGDLSQNRLLYPGDIIVPBNDDLKVFVMEVKRKOSTLKKDFSGM  
 TLEVALGNAEGIDLTTSNAGSIFVIRPLKGGGGRGKIANIYOLDMSDATSLVMADEF  
 RLQPYDVVYVYTPAVARWRLINQLLFTISGVRYMTDFTASDIHSW"  
 complement(2680. .2685)  
 /gene="wzb"  
 /note="putative RBS for wzb; Reguondb: STMS1H002263"  
 /complement(3918. .3926)  
 /gene="wza"  
 /note="putative -10\_signal for wza; Reguondb: STMT1H004524"  
 /complement(3937. .3945)  
 /gene="wza"  
 /note="putative -35\_signal for wza; Reguondb: STMT1H004524"  
 /complement(4476. .4481)  
 /gene="yegH"  
 /note="yegH"  
 /gene="yegH"  
 /note="putative RBS for yegH; Reguondb: STMS1H002264"  
 /complement(4493. .6073)  
 /gene="yegH"  
 /note="Ortholog of E. coli putative transport protein (AAC75124.1); Blastp hit to AAC75124.1 (549 aa), 90% identity in aa 23 - 549"

```
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL21023.1"
/db_xref="GI:16420653"
/translation="MEWIDPSIWAGLVTLVIELVIGIDNLVFIALLAEKLPQGRD
RARITGLIAMIRLLLASISWLVTLKPLFVSQALSADLIMLFGFFLLFKAT
MELNERLEGKDSANPTORAKAFNAVAQIVYVDAIFSDSVITAVGVHIAVMAA
VITISIMLASKSLTRFVNNPTIIVICLSFLMIGSLVAGGFHHPKGYLYAAI
GSVMTEALNOLAFNRRRLSANHTLROTFTEAVMRLLSGKKEDELDAETAMLAD
HDSQITFNPQERRMTERVNLNQRVSSIMTSRDIHIDILNAPAEIRALEKQHT
RLVTVGENEQEDLVGVHVIDLQSLGEPLNLALIPVPEPETILPLPALEOR
NARHFAFVDEFGSVEGIVTLDVMTIAGLNPVEIDARDIQKNPDGSWTANG
HMPLEDVQYVPLPDEKREYHTIAGLMEYLQIPQTEEVQVDYLLKTLQVESHR
VOKVOLIPLHNGEPPQEW"
complement(6159..8024)
/gene="asma"
/locus_tag="STM2120"
/complement(6159..8015)
/gene="asma"
/notes="Ortholog of E. coli suppressor of ompF assembly
mutants (AAC75125.1); Blastp hit to AAC75125.1 (617 aa),
76% identity in aa 1 - 617"
/codon_start=1
/transl_table=11
/product="suppressor of ompF assembly mutants"
/protein_id="AAL21024.1"
/db_xref="GI:16420654"
/translation="MKRFLITLMLLVIVAGFSALVLLVNPDRFAYMVOQVARSSE
YOLQDGLPLRMVHPQSLISGRMTLTAGASEPLVRADNMRLDVALMPLSHQIHKV
QVWLKGVVIOITPQTEAVRSDDAVPAKDNLTLDVAEDGWSFDIRSLVADSVLVFQ
HENDQDVTRDIRLNMEDDAERHGTEDFSGRVNRDQRIALSGSTVADSDYPNLTIA
GTEQLRMQLOGADLPAQGLEGOGLOAQWQEAOKRSLSEFNLNLTFANDSSLTGOVOTL
ARPEWQIDLOSSKILNDLPLPHSAVAQOTGAVGOGONTLPLTRPVIASRYGAPPYO
GIQSTFAETALQADSVRMRGMPFOYSTMSQAGLIDITDIOGKLADGEMSLPSTLD
ARTASPRIEFHRLNHEVETGTLKAFNPISLTGKSLVDFSGADIDAEARHSHKQ
KAHVDMSNTRLEGMFQOLVQOAVERSGDAQSQENMDNATRLDRFTTDLINKGTL
TLDVWVGSSMLALTGSTLDLVKQSCDTPFNLRVYGWNGDSNLITFLKETPVPLRV
YGRWQELNTLOYDQLLRKHLDEAKRRLNDWADRNRKDTNRNGDKVKKLNKL"
complement(8019..8024)
/gene="asma"
/locus_tag="STM2121"
/complement(8054..8649)
/gene="dcd"
/notes="STM2121"
/complement(8054..8635)
/gene="dcd"
/EC_number="3.5.4.13"
/notes="Ortholog of E. coli 2'-deoxycytidine
5'-triphosphate deaminase (AAC75126.1); Blastp hit to
AAC75126.1 (193 aa), 95% identity in aa 1 - 193"
/codon_start=1
/transl_table=11
/product="dUTPase"
```

```
Query Match 67.6%; Score 19.6; DB 1; Length 26591;
Best Local Similarity 84.6%; Pred. No. 9e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 caccgaagcgcgcgtccagccac 26
||||| ||||| ||||| |||||
DB 8980 CACGGAACGCGCGGTGAGTAAC 9005
```

Search completed: July 31, 2002, 14:00:49  
Job time: 17491 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 09:06:17 ; Search time 165.21 Seconds  
(without alignments)  
5627.521 Million cell updates/sec

Title: US-09-899-718a-1

Perfect score: 3785  
Sequence: 1 gtttggttcgctgttttc.....tggaccgctgttcgac 3785

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID                   | Description        |
|------------|-------|-------------|--------|----------------------|--------------------|
| 1          | 185   | 4.9         | 4800   | 3 US-08-941-445A-4   | Sequence 4, Appl1  |
| 2          | 137.8 | 3.6         | 2542   | 3 US-08-941-445A-6   | Sequence 6, Appl1  |
| 3          | 118.8 | 3.1         | 2267   | 4 US-08-679-645-25   | Sequence 25, Appl1 |
| 4          | 83    | 2.2         | 2549   | 1 US-08-470-720-2    | Sequence 2, Appl1  |
| 5          | 83    | 2.2         | 4964   | 1 US-08-470-720-5    | Sequence 5, Appl1  |
| 6          | 71.2  | 1.9         | 19124  | 2 US-08-487-826B-13  | Sequence 13, Appl1 |
| 7          | 64.8  | 1.7         | 7218   | 1 US-08-232-463-14   | Sequence 14, Appl1 |
| 8          | 62.2  | 1.6         | 7218   | 1 US-08-232-463-14   | Sequence 14, Appl1 |
| 9          | 58.8  | 1.6         | 5852   | 1 US-07-867-106-2    | Sequence 2, Appl1  |
| 10         | 56.6  | 1.5         | 658    | 4 US-08-998-416-595  | Sequence 595, App  |
| 11         | 56.4  | 1.5         | 5852   | 1 US-07-867-106-2    | Sequence 2, Appl1  |
| 12         | 55.6  | 1.5         | 51952  | 3 US-08-947-823-1    | Sequence 1, Appl1  |
| 13         | 54.2  | 1.4         | 615    | 4 US-08-998-416-186  | Sequence 186, App  |
| 14         | 53.8  | 1.4         | 636    | 4 US-08-998-416-1137 | Sequence 1137, App |
| 15         | 52.4  | 1.4         | 837    | 4 US-08-998-416-288  | Sequence 288, App  |
| 16         | 52.2  | 1.4         | 19124  | 2 US-08-487-826B-13  | Sequence 13, Appl1 |
| 17         | 51.2  | 1.4         | 240    | 1 US-08-628-417-6    | Sequence 6, Appl1  |
| 18         | 51    | 1.3         | 1511   | 1 US-07-991-867B-8   | Sequence 8, Appl1  |
| 19         | 51    | 1.3         | 1511   | 1 US-08-107-755A-8   | Sequence 8, Appl1  |
| 20         | 51    | 1.3         | 1511   | 2 US-08-544-332-8    | Sequence 8, Appl1  |
| 21         | 50.4  | 1.3         | 4673   | 1 US-07-638-431-1    | Sequence 1, Appl1  |
| 22         | 50.4  | 1.3         | 4673   | 5 PCT-US93-00018-1   | Sequence 1, Appl1  |
| 23         | 50.4  | 1.3         | 4766   | 5 PCT-US93-07261-10  | Sequence 10, Appl1 |
| 24         | 50.2  | 1.3         | 5181   | 1 US-08-257-073-10   | Sequence 10, Appl1 |
| 25         | 49.6  | 1.3         | 3211   | 2 US-08-574-959A-8   | Sequence 8, Appl1  |
| 26         | 49.6  | 1.3         | 3211   | 2 US-09-357-014-8    | Sequence 8, Appl1  |
| 27         | 49.6  | 1.3         | 3901   | 2 US-08-574-959A-6   | Sequence 6, Appl1  |

|    |      |     |       |                     |                    |
|----|------|-----|-------|---------------------|--------------------|
| 28 | 49.6 | 1.3 | 3901  | 4 US-09-357-014-6   | Sequence 6, Appl1  |
| 29 | 49.6 | 1.3 | 71989 | 4 US-09-443-501A-2  | Sequence 2, Appl1  |
| 30 | 49   | 1.3 | 6243  | 2 US-09-056-075-1   | Sequence 1, Appl1  |
| 31 | 48.6 | 1.3 | 2230  | 4 US-08-378-313-24  | Sequence 24, Appl1 |
| 32 | 48.6 | 1.3 | 2447  | 2 US-09-014-969-14  | Sequence 14, Appl1 |
| 33 | 48.6 | 1.3 | 43795 | 3 US-08-742-185-101 | Sequence 101, App  |
| 34 | 48.4 | 1.3 | 4818  | 3 US-08-817-926-27  | Sequence 3, Appl1  |
| 35 | 47.8 | 1.3 | 2997  | 3 US-08-486-343A-3  | Sequence 3, Appl1  |
| 36 | 47.8 | 1.3 | 2997  | 5 PCT-US95-07349-3  | Sequence 3, Appl1  |
| 37 | 47.4 | 1.3 | 5394  | 3 US-08-688-376-1   | Sequence 1, Appl1  |
| 38 | 47.2 | 1.2 | 602   | 1 US-08-764-100-8   | Sequence 8, Appl1  |
| 39 | 47.2 | 1.2 | 642   | 1 US-08-764-100-13  | Sequence 13, Appl1 |
| 40 | 47.2 | 1.2 | 643   | 1 US-08-764-100-7   | Sequence 7, Appl1  |
| 41 | 47.2 | 1.2 | 2993  | 1 US-08-764-100-2   | Sequence 2, Appl1  |
| 42 | 47.2 | 1.2 | 2993  | 1 US-08-764-100-2   | Sequence 2, Appl1  |
| 43 | 47.2 | 1.2 | 3000  | 1 US-08-764-100-9   | Sequence 9, Appl1  |
| 44 | 47.2 | 1.2 | 3001  | 1 US-08-764-100-1   | Sequence 1, Appl1  |
| 45 | 46.6 | 1.2 | 51952 | 3 US-08-947-823-1   | Sequence 1, Appl1  |

#### ALIGNMENTS

RESULT 1  
US-08-941-445A-4  
Sequence 4, Application US/08941445A  
Patent No. 6107060  
GENERAL INFORMATION:  
APPLICANT: Keeling, Peter  
TITLE OF INVENTION: Starch Encapsulation  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
STREET: 5370 Manhattan Circle  
CITY: Boulder  
STATE: CO  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/941,445A  
FILING DATE: 30-SEP-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/026,855  
FILING DATE: 30-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Winner, Ellen P  
REGISTRATION NUMBER: 28,547  
REFERENCE/DOCKET NUMBER: 89-97  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4800 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: not relevant  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(1449..1553, 1685..1765, 1860..1958, 2055  
..2144, 2226..2289, 2413..2513, 2651..2760, 2858  
..3101, 3212..3394, 3490..3681, 3793..3879, 3977





```

QY 3559 ctgggaacccagcgtcatctccgcgggatatatccgcacaatgaattatcaaatccaca 3658
Db 352 TTGGGATACTGCGCGCTGGCGGGTTGAGGTACATCTTCTATTGATACGGTACAAAT---AT 408
QY 3659 tgcctctgcacattctctcaagaacttactgactgctgatactccgcagatcaagtcgt 3718
Db 409 TGTCTCTTACATATTCCCGA-----TTCAGAAATGTGATCATCT -GCAGGTCAAGAGTTGG 462
QY 3719 tgaacgtagcagagaggtgtgagttacttccactgtctacaagcgcggtgtgagacgcgttt 3778
Db 463 AGACAGCAATTAATAATTTGCTTTCTTCTTCACTGCTATTAAACGTGGGGTTGACGCGTGT 522
QY 3779 cctgcac 3785
Db 523 TGTTCAC 529

RESULT 5
US-08-470-720-5
: Sequence 5, Application US/08470720
: Patent No. 5824798
: GENERAL INFORMATION:
: APPLICANT: HOFVANDER, Per
: APPLICANT: PERSON, Per T
: APPLICANT: WIKSTROM, OLE
: APPLICANT: TALBERG, AnneI
: TITLE OF INVENTION: GENETICALLY ENGINEERED MODIFICATION OF
: TITLE OF INVENTION: POTATO T0 FORM AMYLOPECTIN-TYPE STARCH
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Burns, Doane, Swecker & Mathis
: STREET: George Mason Bldg., Washington & Prince Sts.
: CITY: Alexandria
: STATE: Virginia
: COUNTRY: United States
: ZIP: 22313-1404
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/470,720
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/070,455
: FILING DATE: 09-JUN-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Crane-Feury, Sharon E
: REGISTRATION NUMBER: 36,113
: REFERENCE/DOCKET NUMBER: 003300-293
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-6620
: TELEFAX: (703) 836-2021
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4964 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-470-720-5

```

|    |      |                                                                   |      |
|----|------|-------------------------------------------------------------------|------|
| OY | 3425 | gagcaactccctggggccctcccccgcgcacatggccgtaag-cttgcgcacactgcctct     | 3483 |
| Db | 1497 | GGTATGTTCTTGSTGACACTACACACAGCCCTTGCAGTAAGCTTCTTTATTTGGTTA         | 1556 |
| OY | 3484 | tataatglttctctcctgcagcacaatgcctgcgtttacaaaggatgcgtgtc-----cgt     | 3538 |
| Db | 1557 | CCCTACATTCATCTACTTATTATTGTTTAACTAGTTTCTCTACCTGACACAGCTTTTATCAT    | 1616 |
| OY | 3539 | gcagagccaaaggtcacccgggtctcagtcctcccccgcctcagccagctcagaagcgc       | 3598 |
| Db | 1617 | TTAAGCCCGGGAGAGGGGATATGACAAATATCCCCCGTTATGACCAATACAAAGATGC        | 1676 |
| OY | 3599 | ctgggacacccagcgtcatctcccgaglatataatccgcacatgaattatccaaattcaca     | 3658 |
| Db | 1677 | TTGGGATACTGCGCTTTCGGGTTGAGGTACATCTTCTATATTGATGAACGGTACAAT--AT     | 1733 |
| OY | 3659 | tgctctctgcacattctctgcagaacttactgcactgcgtgcagatctcgcagatcaagtcgt   | 3718 |
| Db | 1734 | TGTTCTCTTACATTTCTCTGA-----TTCAAGAGATGTGATCATCT-GCAGGTCAAAGTGG     | 1787 |
| OY | 3719 | tgcaagtgatcgaagaggtgtgagttacttccactgcatacgaagcgcgggggtggagccggtgt | 3778 |
| Db | 1788 | AGACAGCATTCACAAATTTGCTGTTCTTTCACGTCATTAACCGTGGGGTGATCGTGT         | 1847 |
| OY | 3779 | cgtcgac                                                           | 3785 |
| Db | 1848 | TGTTGAC                                                           | 1854 |

```

1  RESULT 6
2  US-08-487-826B-13/c
3  : Sequence 13, Application US/08487826B
4  : Patent No. 5993827
5  :
6  : GENERAL INFORMATION:
7  :
8  : APPLICANT: Sim, Kim L.
9  : APPLICANT: Chitnis, Chetan
10 : APPLICANT: Miller, Louis H.
11 : APPLICANT: Peterson, David S.
12 : APPLICANT: Su, Xin-zhaun
13 : APPLICANT: Wellens, Thomas E.
14 :
15 : TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
16 : TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
17 :
18 : NUMBER OF SEQUENCES: 45
19 :
20 : CORRESPONDENCE ADDRESS:
21 :
22 : ADDRESSEE: Knobbe Martens Olson & Bear
23 : STREET: 620 Newport Center Drive 16th Floor
24 : CITY: Newport Beach
25 : STATE: California
26 : COUNTRY: US
27 :
28 : ZIP: 92660
29 :
30 : COMPUTER READABLE FORM:
31 :
32 : MEDIUM TYPE: Floppy disk
33 : COMPUTER: IBM PC compatible
34 : OPERATING SYSTEM: PC-DOS/MS-DOS
35 : SOFTWARE: PatentIn Release #1.0, Version #1.25
36 :
37 : CURRENT APPLICATION DATA:
38 :
39 : APPLICATION NUMBER: US/08/487, 826B
40 : FILING DATE: 10-SEP-1993
41 : CLASSIFICATION: 435
42 :
43 : ATTORNEY/AGENT INFORMATION:
44 :
45 : NAME: Israelsen, Ned
46 : REGISTRATION NUMBER: 29,655
47 : REFERENCE/DOCKET NUMBER: NIH121.001CP1
48 :
49 : TELECOMMUNICATION INFORMATION:
50 :
51 : TELEPHONE: (619) 235-8550
52 : TELEFAX: (619) 235-0176
53 :
54 : INFORMATION FOR SEQ ID NO: 13:
55 :
56 : SEQUENCE CHARACTERISTICS:
57 :
58 : LENGTH: 19124 base pairs
59 : TYPE: nucleic acid
60 : STRANDEDNESS: single
61 : TOPOLOGY: linear
62 :
63 : MOLECULE TYPE: cDNA
64 :

```







```

Db      1832 TTTTCTTTTTTTTTTTTCCTTTTTTTTTTTAAATTTTAAATTTTTTTTTTTTTTAATAGATGTCA 189
QY      453 tttagaaattcgttttggaaaataatgctgtaacaaacttatctgtatatattgaga 512
Db      1892 TAATTAAAATAACCAATTAATAAATTAATAAGTTATTTTAAATATGCAAAAACATAAAAAAAC 1951
QY      513 agcgcaattctgtgaagtgttgcatctgtatcttatttcattcttctctctcg 572
Db      1952 TAAGTAGACTTTAACCACTTTTTCATATTCTTTTTTTTTTTTTTTTTTTTTTTTACTTT 2011
QY      573 -----aagggtaacacaatgcacaaattcatctgtctaganaaatctgattcttga 628
Db      2012 GAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACCCTCATTAATTAATTAATTAATTAATGG 2071
QY      629 ttgtgttttagtttttttattctattctgtttctctctctttaaagggaataacaaagcacaa 688
Db      2072 TTTTCTTTTGATTTTTTTTTTTAAATAAATTTAAATTTTAAATTTCTCCTATTAATTAATTAACGCTTA 2131
QY      689 tccattccattctagaaaatctcttattcttcaaaaaaacaaactttaalgcctatlc 748
Db      2132 TTFA-TAAATATTTGATTAATATATATCAAAATTTTATCATCGTTTGGCATGCATTTTAAAT 2190
QY      749 gtgcattattlaaaaagcacagtttctatctaattcggtgcaaccttatcatattg 808
Db      2191 AATTTAATTTTTTGATTAATTAATTTTTTTTTTTTTTTTTTTTTTTTTTAAATTTCTTTTTTTTTTTT 2250
QY      809 tcataattaatlctt 824
Db      2251 TTTAATTTTAAATTTT 2266

RESULT 10
US-08-998-416-595
; Sequence 595, Application US/0898416
; Patent NO. 6239264
;
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jürgen
APPLICANT: Knechtle, Philipp
APPLICANT: Repischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPIT
TITLE OF INVENTION: AND USICS THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264Ch Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38, 241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8567
TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 595:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 658 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAGI408BP
US-08-998-416-595

Query Match 1.5%; Score 56.6; DB 4; Length 658;
Best Local Similarity 47.0%; Pred. No. 0.00035;
Matches 209; Conservative 0; Mismatches 234; Indels 2; Gaps 1

Oy 203 ttcttacagttcagtttattatctgtttattaaagtglttttagtggtttctcatt 262
Db ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
216 ttttgattattttatttttttttttttttttttttttttttttttttttttttttt 275
Oy 263 tatgcatcagcatgaataatlaagggtgtgtgtgtgtgtgtgttaatacacaaagatta 322
Db 276 attatttttttttttttttttttttttttttttttttttttttttttttttttt 335
Oy 323 tacaccacatttttcagtcataaataatgcaatttcaglacaaattgtgcgcaacct 382
Db 336 tttatttttttttttttttttttttttttttttttttttttttttttttttttttt 395
Oy 383 tcttcatttttttttttttttttttttttttttttttttttttttttttttttttt 442
Db 396 tttttatttttttttttttttttttttttttttttttttttttttttttttttttt 455
Oy 443 ttatgctcattttgaaatttcgttttgaaataatgctagcacacacttattctgtat 502
Db 456 tcatatgattatgatt--atattatgctgacttttttttttttttttttttttttttt 513
Oy 503 attaggaagaagcgcaatttcgtgtgaagtttgcattctgtattttttcattttc 562
Db 514 ataatatattatucgattttctattttctatttttttttttttttttttttttttttt 573
Oy 563 ttctctctggaagglaacacactaatgcccactaatcattctgcgtctagaagaacttagta 622
Db 574 tattttttttttttttttttttttttttttttttttttttttttttttttttttttt 633
Oy 623 tttgatctgtcttttagtttttatt 647
Db 634 tattttttttttttttttttttttttttttttttttttttttttttttttttttttt 658

RESULT 11
US-07-867-106-2/c
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kuriz Mackiewicz & No. 53895261ris
; STREET: One Liberty Place 46th Floor
; City: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/867,106

```

FILING DATE: 19920625  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PJ 7187  
APPLICATION NUMBER: PCT/AU90/00530  
FILING DATE: 02-NOV-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Feeney, Joanne Longo  
REGISTRATION NUMBER: 35,134  
REFERENCE/DOCKET NUMBER: RICE-0002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-568-3100  
TELEFAX: 215-568-3439  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5852 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2378..5038  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2378..5038  
US-07-867-106-2

Query Match 1.5%; Score 56.4; DB 1; Length 5852;  
Best Local Similarity 48.1%; Pred. No. 0.0011;  
Matches 228; Conservative 0; Mismatches 236; Indels 10; Gaps 2;

OY 314 taagtattatccaccatttttgcagtcataaattatgcatttcgtatcaatttgc 373  
DB 5549 TTAATAATTATGATCGATTTTAAATCAATGACGATTAATAAATCTTAAAC 5490  
OY 374 gcaactcttcatttatttatttatttcttcttaagggtatcaaat 433  
DB 5489 AATAAACATATTGATTTTATTTTATTTTATTTTAAATCAATATATTA 5430  
OY 434 gatactaattatgcatttggaatttcgttttggaattatgctag------ta 485  
DB 5429 AATAATATATATATATATATATATATATATATATATATATATATATAT 5370  
OY 486 cacactttcttgatattatggaagaagcatttcctgtatgatttgctatctgt 545  
DB 5369 AAAAATATGATATGATATGATATATATATATATATATATATATATATAT 5310  
OY 546 atttttttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 605  
DB 5309 ATTATTTAATTAATAGACGACGATTAATATGAAATGAAATGAAATGAAATG 5250  
OY 606 cttagaataacttagtatttgatttgatttgatttgatttgatttgatttgatt 665  
DB 5249 CCAATTTTCCCTTATATATATATATATATATATATATATATATATATAT 5192  
OY 666 aagggaataacacacacacacacacacacacacacacacacacacacacacac 725  
DB 5191 CCAGGCAAAATTAACAGATCGAATCTCATTAATATTTGCTCTCAAAATTTT 5132  
OY 726 actcaacttatgcttatgcttatgcttatgcttatgcttatgcttatgcttat 779  
DB 5131 TCTCATTTATATATATATATATATATATATATATATATATATATATATAT 5078

RESULT 12  
US-08-947-823-1  
Sequence 1, Application US/08947823  
Patent No. 6114605  
GENERAL INFORMATION:  
APPLICANT: Williamson, Valerie M.  
APPLICANT: Kaloshian, Isouhi

APPLICANT: Yaghoobi, Jafar  
APPLICANT: Bodeau, John  
APPLICANT: Milligan, Stephen  
TITLE OF INVENTION: Procedures and Materials for Conferring  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESS: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/947,823  
FILING DATE: 09-OCT-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/18802  
FILING DATE: 09-OCT-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/028,191  
FILING DATE: 10-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 023070-07021005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 51952 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-947-823-1

Query Match 1.5%; Score 55.6; DB 3; Length 51952;  
Best Local Similarity 43.9%; Pred. No. 0.005;  
Matches 238; Conservative 0; Mismatches 304; Indels 0; Gaps 0;

OY 125 tcatgtttatttatttcttcttcttcttcttcttcttcttcttcttcttcttct 184  
DB 12960 TAATGTCCTCTTATATATATATATATATATATATATATATATATATAT 13019  
OY 185 acttaagggaattttagtttatttacttagtttcaagtttatttatttatttatt 244  
DB 13020 TGCTTATATATATATATATATATATATATATATATATATATATATAT 13079  
OY 245 tttagttggtttcttcttcttcttcttcttcttcttcttcttcttcttcttct 304  
DB 13080 TTTTTCCTTATATATATATATATATATATATATATATATATATATATAT 13139  
OY 305 atatacaataaglatataacacacacacacacacacacacacacacacacac 364  
DB 13140 TACTCTTAATATATATATATATATATATATATATATATATATATATAT 13159  
OY 365 aaatttggtgcaacacacacacacacacacacacacacacacacacacacac 424  
DB 13200 AAACCTTAAGATATATATATATATATATATATATATATATATATATAT 13259  
OY 425 aatacaatgataactatttatttccatttgggaatttcgatttggaatttgc 484  
DB 13260 AACCCCTACCAAGATATATATATATATATATATATATATATATATATAT 13319



TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 1137:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 636 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAGI692RP  
US-09-998-416-1137

Query Match 1.48; Score 53.8; DB 4; Length 636;  
Best Local Similarity 45.98; Pred. No. 0.0016;  
Matches 266; Conservative 0; Mismatches 302; Indels 11; Gaps 2;

```
QY 196 aattcagtttatactagttcagtttatactgatttataagtggtttggttt 255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 58 ATTAGTATTAATAATTATTAACATATTAATCAATTAATTAATAATTAATTAAT 117
QY 256 tctcattatgtagtcaatgaaatgaggggtggtggtggtggtggtggtggt 315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 TATTAACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 177
QY 316 agtatatacaccatttgcagtcataaataagcatttcagtcagtcagtcagtc 375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 TATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 237
QY 376 aaactctctcatttatttatttatttatttatttatttatttatttatttattt 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 TAGTATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 297
QY 436 tactaattatgctcatttgcgaattcgtttgaa-----ttatgctagtaga 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 CAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 357
QY 488 caattatctgataatattgaaagcgaattctgtagt-----ttgtagctctg 544
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 GATTTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 417
QY 545 tatttttttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 604
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 418 TATTAGTGAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 477
QY 605 gcttagaagaactttagtattttagtattttagtattttagtattttagtattt 664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 478 ACTTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 537
QY 665 taaggaaatcaccaatgcacaaatccatccatccatccatccatccatccatcc 724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 538 TTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 597
QY 725 aactcaactttatagcttattcgtgcataattataaa 763
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 598 AAAATATTAATAAGATGAGTTAAATAATAATAATAATAATAATAATAATAATA 636
```

RESULT 15  
US-09-998-416-288  
Sequence 288, Application US/08998416  
Patent No. 6239264

## GENERAL INFORMATION:

APPLICANT: Philippsen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jurgen  
APPLICANT: Knechtel, Philipp  
APPLICANT: Reibschung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPYII  
AND USES THEREOF

NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/GC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 288:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 837 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1241RP  
US-09-998-416-288

Query Match 1.48; Score 52.4; DB 4; Length 837;  
Best Local Similarity 44.68; Pred. No. 0.0039;  
Matches 258; Conservative 0; Mismatches 311; Indels 9; Gaps 1;

```
QY 196 aattcagtttatactagttcagtttatactgatttataagtggtttggttt 255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 58 ATTAGTATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 117
QY 256 tctcattatgtagtcaatgaaatgaggggtggtggtggtggtggtggtggtggt 315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 TATTAACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 177
QY 316 agtatatacaccatttgcagtcataaataagcatttcagtcagtcagtcagtcagtc 375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 TATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 237
QY 376 aaactctctcatttatttatttatttatttatttatttatttatttatttattt 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 TAGTATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 297
QY 436 tactaattatgctcatttgcgaattcgtttgaa-----ttatgctagtaga 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 CAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 357
QY 487 aactcattctgataatgaaagcgaattctgtagtattttagtattttagtattt 546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 GATTTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 417
QY 547 ttttttttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 606
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 418 TATTAGTGAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 477
QY 607 ttagaagaactttagtattttagtattttagtattttagtattttagtattttagt 666
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 478 ACTTTAAATTTTCCTTATTTATTTATTTTAAATTTAAATTAATTTAATTCATTTTATTA 537

QY 667 agggaaatacaacatggccactatcatccatcctgagaaaatccttactcttaacaaaa 726

Db 538 TTTATTTATTTTAAATTAATTAATTTATTTAATTTATTTATTTATTTATTTAATTAATTA 587

QY 727 ctcaactttatatacgctatccgycgatatactataaaaa 764

Db 598 TAAATTTATTTATTAAGATAGTGTAAATTAATTTACTTATTA 635

Search completed: July 31, 2002, 12:17:49  
Job time: 11492 sec











Page

AL508163      700 bp      mRNA      linear      EST 04-JAN-2001  
 AL508163      Hordeum vulgare      Barke developing      carypsps (3--15 DAP),  
 AL508163      Hordeum vulgare      cDNA clone H107P10V 5',      mRNA sequence  
 AL508163.1      GI:12034370      EST

*barley*,  
*Hordeum vulgare*  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta;  
Tricicolae; Magnoliophyta; Liliopsida; Poales; Pooideae; Poifiteae;  
(base 1 to 700)  
Schleicher, W.; Meschke, W.

published online first in the *Journal of Plant Physiology* 2006, 163, 1111-1116.  
 doi:10.1016/j.jplph.2006.05.005  
 Copyright © 2006 Elsevier GmbH. All rights reserved.  
 Correspondence: Dr. Michael W. Gatersleben, Institute for Plant Genetics and Crop Plant Research, P.O. Box 607081, 06100 Halle, Germany.  
 E-mail: michael.w.gatersleben@ipk-gatersleben.de  
 Received 11 May 2006; accepted 11 May 2006

```

/clone="Hy07p10v"
/clone_11b="Hordium vulgare Barke developing caryopsis (3-15 DAP)"
/lisuse_type="developing caryopsis (3-15 DAP)"
/lab_host="XLOLR"
/notes="Vector: pVec1"

```

Query Match  
 BASE COUNT  
 ORIGIN  
 133 a 245 c 213 g 107 t 2 others  
 'various barley varieties from developing site-1: EcoRI; Site-2:  
 (3'-end of cDNA). NOTE: EcoRI (5'-end of cDNA) and XhoI  
 by the kit, in most cases the EcoRI adapter and XhoI  
 Sequence as the EcoRI adapter. Average insert size is  
 trimmed. trimming: Vector adapter. Average insert size is  
 contained from the 5'- and 3'-end until a 50 bp window  
 containing less than two ambiguities. The maximum length was

[illegible]

QY 3472 cc 3473  
 ||  
 Db 530 CC 531

RESULT 7  
 AL508933  
 LOCUS 513 bp mRNA linear EST 04-JAN-2001  
 DEFINITION Hordeum vulgare Barke developing caryopsis (3.-15.DAP)  
 AL508933  
 Hordeum vulgare cDNA clone HY10D16V 5', mRNA sequence.  
 ACCESSION  
 AL508933  
 VERSION  
 EST.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Barley.  
 Hordeum vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae  
 ; Triticeae; Hordeum.  
 1 (bases 1 to 513)  
 ; Triliceae; Hordeum.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Institute for Plant Genetics and Crop Plant Research  
 Corrensstr.3, D-06466 Gatersleben, Germany  
 Email: michael@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de  
 Seq primer: T3 primer for 5' end.  
 Location/Qualifiers  
 1. 513  
 /organism="Hordeum vulgare"  
 /cultivar="Barke"  
 /db\_xref="taxon:4513"  
 /clone\_11b="HY10D16V"  
 /clone\_11b="Hordeum vulgare Barke developing caryopsis  
 (3.-15.DAP)"  
 /tissue\_type="developing caryopsis (3.-15.DAP)"  
 /lab\_host="XLOLR"  
 /note="Vector: Plasmid pBK-CMV; Site\_1: EcoRI; Site\_2:  
 XhoI; mRNA was made from developing caryopsis (3.-15.DAP)  
 of spring barley variety 'Barke', a high quality malting  
 variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI  
 (3'-end of cDNA). NOTE: Due to a cloning artefact caused  
 by the kit, in most cases the EcoRI site is NOT present,  
 as well as the EcoRI adapter. Average insert size is 1 kb  
 trimmed from the 5'- and 3'-end until a 50 bp window  
 contains less than two ambiguities. The maximum length was  
 set to 700 bp"

BASE COUNT 102 a 184 c 151 g 73 t 3 others  
 ORIGIN

Query Match 6.8%; Score 258.8; DB 9; Length 513;  
 Best Local Similarity 90.9%; Pred. No. 3.1e-28;  
 Matches 298; Conservative 0; Mismatches 24; Indels 6; Gaps 2;

QY 3099 cagtcgtctctctgtagcaacacccctgcgc--gcgcacatgcygctctgtgca 3155  
 |||  
 Db 188 CAGTCGCTCTCTACGTGAGGTACACACCTGTGCGGCGCCGATGGCGGCTCTGCGCA 247

QY 3156 cgtccacagctcgcacactcgcacgctcctcagctaacccgaattccggcgctcag 3215  
 |||  
 Db 248 CGTCCAGCTCGCACCTCCGACACCTCTCGGNGTCAACCGACATTCGCGCTCCAG 307

QY 3216 gtttcaaggcctgaagcccggaacccgagatgcgcgctcgcgacatgaagacgtcg 3275  
 |||  
 Db 308 GTTTTCAAGGCTCTCAGGCCCCGGAACCGAGCGGATGCGGCTGTATGAGGACTATCG 367

QY 3276 gaacgagcgccgcccccaagaagcaagaaacgcacgcatgcacgcgagtgctct 3335  
 |||  
 Db 368 GACCAAGCGCCGNCGCCGAAGCAAAAGCCGACCGCGGAGACCGGCGGTGCTCT 427

QY 3336 ccattggtggtggtgcgcgcacagcgagcgagtgactgctgtctgcgcgcgaga 3395  
 |||  
 Db 428 CCGTGTGTGTGCGCGCCACGCGCA--GCGCATGAACCTGTCTGCGCGCCGAGA 484

QY 3396 tggcgcccttggaagaagactgcgcgct 3423  
 |||  
 Db 485 TGGCGCCCTGTGAGCAAGACCGCGGCGCT 512

RESULT 8  
 BE414500  
 LOCUS 677 bp mRNA linear EST 24-JUL-2000  
 DEFINITION SCU010.D11.R990714 TREC SCU Wheat Endosperm Library Triticum  
 aestivum cDNA clone SCU010.D11, mRNA sequence.  
 ACCESSION  
 BE414500  
 VERSION  
 EST.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae  
 ; Triticeae; Triticum.  
 1 (bases 1 to 677)

REFERENCE  
 AUTHORS  
 Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier  
 S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,  
 Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,  
 Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,T.,  
 Pecchioni,N., Qaiset,C., Schuch,W., Selvaraj,G., Shariflou,M.,  
 Sorrells,M., Warburton,M. and Wenzel,G.  
 International Triticaceae EST Cooperative (TREC): Production of  
 Expressed Sequence Tags for Species of the Triticeae  
 Unpublished (2000)

JOURNAL  
 COMMENT  
 Centre for Plant Conservation Genetics, Southern Cross University  
 PO Box 157, Lismore NSW 2480 AUSTRALIA  
 Tel: 61 2 6620 3409  
 Fax: 61 2 6622 2080  
 Email: tholton@scu.edu.au  
 International Triticaceae EST Cooperative (TREC)  
 http://wheat.pw.usda.gov/genome.  
 Location/Qualifiers  
 1. 677  
 /organism="Triticum aestivum"  
 /cultivar="Wynna"  
 /db\_xref="taxon:4565"  
 /clone\_11b="TREC SCU Wheat Endosperm Library"  
 /tissue\_type="endosperm"  
 /note="Vector: Bluescript II SK(-)"

BASE COUNT 138 a 207 c 212 g 97 t 23 others  
 ORIGIN

Query Match 6.6%; Score 248; DB 10; Length 677;  
 Best Local Similarity 83.5%; Pred. No. 1.1e-26;  
 Matches 313; Conservative 0; Mismatches 57; Indels 5; Gaps 3;

QY 3092 atcattgagtgctgtctgtctgtagcaacacccctgcgcgacatgagcgctcg 3151  
 |||  
 Db 254 ATCAACAAGTGTCTCTGCTGCGAGTAGCACACCTGTGCGCGCGGATGGCGCTCTG 313

QY 3152 gtcaagctccagctcgcacactcgcgcacgctcctcagctcagctcagcagatccgct 3211  
 |||  
 Db 314 GTACAGTGTGACAGCTCGCACCTCGGACGCTCTCTCGGACATCACGACGATTCGCGGT 373

QY 3212 ccaggtttcagagcctgtaggcccccggaacccgagatgcgcgctcgcgcatgagact 3271  
 |||  
 Db 374 GCAAGTTTTCAGGCTGTGAGGCCCCGAGCCCGACATGCGCGCTCGCATGAGACT 433

QY 3272 gtccgagagagcgccgccca--aagcaagcagagaaacgcacgcatccgacgcg 3328  
 |||  
 Db 434 ACCGAGAGGAGCGCGCCCGCGAAGCAACAAAGCCCGAAGAGCCCGCGGAGACCGCGCG 493







```
/organism="Tetraodon nigroviridis"  
/db_xref="taxon:99883"  
/clone="208p24"  
/clone_1tb="G"
```

|    |     |                                                              |     |
|----|-----|--------------------------------------------------------------|-----|
| QY | 2   | tttgatctgcgcgttttcaattccttctctcttaaggaggtaaaccaatgcagtaatt   | 61  |
| Db | 969 | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGTTAAAGATATTTTATTTTGT      | 910 |
| QY | 62  | catatctgtaacagtcgcatctctgtgcgaattatgacaattctctctgttaatt      | 122 |
| Db | 909 | TTTGGTTTTTTTAAATGATTTATTTTATTTTATTTTNTTTTTTTTTTTTTTTT        | 850 |
| QY | 122 | gttcaatgtttattcattcttcttcaattcttagaggtaaaccaatgcccaattcat    | 181 |
| Db | 849 | TTTTTTTTTTTTTTTTTTTTTTRGCGTTTGTGTTTATTTTATTTATTTATTTATTT     | 790 |
| QY | 182 | tcacaactaaggagaatccagttctataacagattcagttatatagtttataagt      | 24  |
| Db | 789 | TTTTTTTTTA-WTTGTGTTTTGCAKTTTTTTTTTTTTTTTTTANANAGTAAATTTTTT   | 731 |
| QY | 242 | tttttagtctggtttctccattatgtaagatgaataatlaggggtgtgtgcgtg       | 302 |
| Db | 730 | TTTTTTTTTTTTTATATTTTTTTTTTTTTTTTTTTTTTATTTAAATTTTTTTTTTAAATA | 672 |
| QY | 302 | ttaataacacaataagatgataacaccatttttgcagacataaataatgcgaattccag  | 363 |
| Db | 670 | TTTATTTTTTTTTTAAATTTTTTATTTTTTTTAAATTTTTTTTTTTTTTTTTTAAATTTT | 611 |
| QY | 362 | tacaactgtgcgaactctctctcaattttattttatttatttattcttcttaag       | 421 |
| Db | 610 | TTTTTTTTTTTTTTTTTATTTTATTTTTTTTTTTTTTTTTTTTTTTTATTTATTTT     | 551 |
| QY | 422 | ggtaataacaaagatacctaatttagtcctcaattggaaattccgtttgaaataatgc   | 481 |
| Db | 550 | TTTTTTTTTAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAATTTTTT      | 490 |
| QY | 482 | agtaacactatctctgatataatgaaaagcgcaattctctgtgaagtttgcaat       | 541 |
| Db | 490 | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTT   | 431 |
| QY | 542 | cgtatatttttcaatttctctctctctcggaaggtaacactaatgccaattcaat      | 601 |
| Db | 430 | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT    | 371 |
| QY | 602 | cttgcttagaaaacttagcatcttgatctgtgttttagtttatttcaattgtctct     | 661 |
| Db | 370 | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT    | 311 |
| QY | 662 | cttt 665                                                     |     |
| Db | 310 | TTTT 307                                                     |     |

Search completed: July 31, 2002, 12:11:51  
Job time: 13464 sec

Thu Aug 1 08:29:57 2002

us-09-899-718a-1.1st

Page 10









```

QY 3241 cccgagc-----gatcgagcgtcgcaatgaactctcgagcgagcgcccccacaa 3294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1381 ccccgccgagcgagcgagcgtcgctcagctgaacgagcgagcgagcgagcgccaa 1440
QY 3295 gaaagcaggaacccagcagcgtcgagcgagcgagcgagcgagcgagcgagcgagc 3354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1441 gaaagcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 1500
QY 3355 gggcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 3414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1501 ---cgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 1557
QY 3415 tggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc----- 3466
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1558 cggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 1617
QY 3467 -----ttcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 3509
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1618 actcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 1677
QY 3510 cctcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 3563
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1678 gggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 1737
QY 3564 gtcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 3623
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1738 gtcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 1797
QY 3624 gta-----atcgagcgagcgagcgagcgagcgagcgagcgagcgagc 3660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1798 gtcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 1857
QY 3661 ctcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 3720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1858 attcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 1917
QY 3721 acagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 3780
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1918 acagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 1977
QY 3781 tcgagc 3785
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1978 tcgagc 1982

```

RESULT 3  
AAV29753 standard; DNA; 2542 BP.

```

XX AC AAV29753;
XX DT 11-SEP-1998 (first entry)
XX DE Oryza sativa waxy gene.
XX KW SBR: starch-encapsulating region; fusion vector;
XX starch synthase; bacterial glycoen; ss.
XX OS Oryza sativa.
XX FH Key
XX FM CDS
XX FT Location/Qualifiers
XX FT 453..2282
XX FT /*tag= a
XX FT /product= starch (bacterial glycoen) synthase
XX FT 2535
XX FT /*tag= b
XX PN WO9814601-A1.
XX PD 09-APR-1998.
XX PP 30-SEP-1997; 97WO-US17555.

```

```

XX PR 30-SEP-1996; 96US-0026855.
XX PA (EXSE-) EXSEED GENETICS LLC.
XX PI Guan H, Keeling P;
XX DR WPI: 1998-240100/21.
XX DR P-PSDB; AAM56485.
XX PT Hybrid polypeptide comprising starch-encapsulating region and
XX protein - useful for, e.g. producing protein(s) resistant to
XX degradation by stomach acids
XX PS Example 2; Page 32-34; 156pp; English.
XX CC The sequence is that of the waxy gene which codes for starch
XX CC synthase. It can be used in the production of a hybrid
XX CC polypeptide comprising a starch-encapsulating region (SBR) fused
XX CC to a payload protein. The hybrid polypeptide can be used to make
XX CC modified starches comprising the payload protein, selected from,
XX CC e.g. hormones, growth factors, antibodies, enzymes, dyes,
XX CC immunoglobulins, etc. The modified starch can also be used
XX CC to provide grain feeds enriched in amino acids. By encapsulating
XX CC the payload protein in starch, it is more resistant to
XX CC degradation by stomach acids.
XX SQ Sequence 2542 BP; 610 A; 665 C; 693 G; 574 T; 0 other;

```

Query Match 3.6%; Score 137.8; DB 19; Length 2542.  
Best Local Similarity 69.5%; Pred. No. 2.7e-21;  
Matches 241; Conservative 0; Mismatches 82; Indels 24; Gaps 3;

```

QY 3136 cgcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 3195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 449 caccatgctcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 508
QY 3196 cgcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 3240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 509 cgcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 568
QY 3241 cccgagc-----gatcgagcgagcgagcgagcgagcgagcgagcgagc 3294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 569 ccccgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 628
QY 3295 gaaagcaggaacccagcagcgtcgagcgagcgagcgagcgagcgagcgagc 3354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 629 gaaagcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 688
QY 3355 gggcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 3414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 689 ---cgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 745
QY 3415 tggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 3461
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 746 cggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 792

```

RESULT 4  
AAV60319 standard; DNA; 1915 BP.

```

XX AC AAV60319;
XX DT 23-AUG-1999 (first entry)
XX DE DNA sequence of the maize waxy gene.
XX KW Non-glycogen-like polysaccharide production; fermentation; waxy gene;
XX starch synthesis enzyme; glycogen-synthesis enzyme; glycogen synthase;
XX non-starch branching gene; amylopectin; amylose; plant-like starch; ss.
XX PP

```

|    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| OS | Zea mays.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| PN | WO9844780-A1.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| XX |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| PD | 15-OCT-1998.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| XX |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| PF | 03-APR-1998; 98WO-US065660.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| XX |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| PR | 04-APR-1997; 97US-0042939.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| XX |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| PA | (EXSE-) EXSEED GENETICS LLC.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| XX |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| PI | Guan H, Keeling PJ;                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| XX |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| DR | WPI: 1998-568285/48.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| DR | P-PSDB; AAY16604.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| XX |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| PT | Producing non-glycogen-like polysaccharides in bacteria, fungi or plants - transformed with genes for enzymes involved in starch or glycogen synthesis allows fermentative production of starches with engineered properties                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| PS | Disclosure; Fig 49; 150pp; English.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| XX |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| CC | The specification describes a method for the production of non-glycogen-like polysaccharides in a host. The method comprises transforming a host, suitable for fermentation, with genes encoding starch- or glycogen-synthesis enzymes, and fermenting the transformants. The specification also describes hosts transformed with a gene active in glycogen synthesis and at least one non-starch branching gene. CC involved in production of amylopectin or amylose in its original host. CC The method is used to produce plant-like starches by fermentation and new starches in plants. These starches are useful for all food and non-food applications of starch. The present sequence is used in the course of the invention. |
| XX |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| SO | Sequence 1915 BP; 365 A; 610 C; 641 G; 299 T; 0 other;                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |

| Query Match  | Best Local Similarity                                                   | 3.3%; | Score 124.4;    | DB 19;    | Length 1915; |
|--------------|-------------------------------------------------------------------------|-------|-----------------|-----------|--------------|
| Matches 203; | Conservative                                                            | 0;    | Mismatches 131; | Indels 0; | Gaps 0;      |
| QY 3140      | atggcgcgtctctgtgtcaccgctccagcctcgcacatctccgcgcgtctccatcaggttcacccgac    |       |                 |           |              |
| Db 1         | atggcgcgtctctgtgtcaccgctccagcctcgcacatctccgcgcgtctccatcaggttcacccgac    |       |                 |           |              |
| QY 3200      | agatctccgcgcgttcacaggtttttcaaggcctctgaaagccaccgaaacccgcgcgatactgcgcctc  |       |                 |           |              |
| Db 61        | gcgtccaccgcgtctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc   |       |                 |           |              |
| QY 3260      | gcgtctgtaggcctctctgtagacagagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc   |       |                 |           |              |
| Db 121       | gcgcgcacccgcgtccagcaccctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc   |       |                 |           |              |
| QY 3320      | gaccgcgcgcgcgtccctccaccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc |       |                 |           |              |
| Db 181       | cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc   |       |                 |           |              |
| QY 3380      | ttcgtctgtagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc  |       |                 |           |              |
| Db 241       | ttcgtctgtagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc  |       |                 |           |              |
| QY 3440      | ggcctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc |       |                 |           |              |
| Db 301       | ggcctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc |       |                 |           |              |

| RESULT   | 5                               |
|----------|---------------------------------|
| AAx63355 |                                 |
| ID       | AAx63355 standard; cDNA; 267 BP |
| XX       |                                 |
| AC       | AAx63355;                       |

|    |                                                                          |
|----|--------------------------------------------------------------------------|
| XX |                                                                          |
| DT | 16-JUL-1999 (first entry)                                                |
| XX |                                                                          |
| DE | Granule bound starch synthase encoding cDNA.                             |
| XX |                                                                          |
| KW | Maize; corn; Zea mays; delta-9 desaturase; GBSS; target; substrate;      |
| KW | granule bound starch synthase; hammerhead ribozyme; hairpin ribozyme;    |
| KW | modulation; gene expression; transgenic plant; cleavage; canola plant;   |
| KW | caffeine synthesis; coffee plant; nicotine production; tobacco;          |
| KW | fruit ripening; flower pigmentation; lignin production; ss.              |
| XX |                                                                          |
| OS | Zea mays.                                                                |
| XX |                                                                          |
| PN | WO9710328-A2.                                                            |
| XX |                                                                          |
| XX | 20-MAR-1997.                                                             |
| XX |                                                                          |
| PF | 12-JUL-1996; 96WO-US11689.                                               |
| XX |                                                                          |
| PR | 13-JUL-1995; 95US-0001135.                                               |
| XX |                                                                          |
| PA | (DOMC ) DOWELANCO.                                                       |
| PA | (RIBO-) RIBOZYME PHARM INC.                                              |
| XX |                                                                          |
| PI | Edgington BE, Folkerts O, Guo L, McSwiggen JA, Merlo DJ;                 |
| PI | Merlo PAO, Skokut TA, Young SA, Zwick MG;                                |
| XX |                                                                          |
| DR | WPI: 1997-202224/18.                                                     |
| XX |                                                                          |
| PT | Ribozyme which modulates plant gene expression - preferably              |
| PT | modulates expression of DELTA-9 desaturase or granule bound starch       |
| PT | synthase in maize or canola                                              |
| XX |                                                                          |
| PS | Example 9; Page 31-33; 155pp; English.                                   |
| XX |                                                                          |
| CC | The present invention describes an enzymatic nucleic acid molecule (I)   |
| CC | with RNA cleaving activity, which modulates the expression of a plant    |
| CC | gene. Also described is a gene comprising a cDNA sequence encoding maize |
| CC | Delta-9 desaturase. (I) can be used to modulate expression of a gene,    |
| CC | preferably Delta-9 desaturase or a granule bound starch synthase (GBSS)  |
| CC | gene, in a plant (preferably a maize or canola plant). (I) can be used   |
| CC | to modulate caffeine synthesis in a coffee plant, nicotine production in |
| CC | a tobacco plant, fruit ripening processes in an apple, tomato, pear,     |
| CC | plum or peach plant, flower pigmentation in a rose, petunia,             |
| CC | chrysanthemum or marigold plant or lignin production in a tobacco,       |
| CC | aspen, poplar or pine plant.                                             |
| XX |                                                                          |
| SO | Sequence 2267 BP; 465 A; 699 C; 733 G; 370 T; 0 other;                   |

[illegible]



PA (RRPG-) RPR GENCELL ASIA PACIFIC INC.  
 XX Hamada H;  
 XX WPI: 1999-243728/20.  
 DR New apoptosis-resistant virus-sensitive cell  
 XX  
 PS Example 2: Page 41-45; 51pp; English.  
 XX  
 CC The present invention describes an apoptosis-resistant virus-sensitive  
 CC cell line into which an apoptosis resistance gene has been introduced.  
 CC The recombinant viruses generated are capable of expressing apoptosis-  
 CC associated genes. These can then be used in a variety of diseases for  
 CC which the induction of apoptosis by gene transfer, or where the  
 CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses  
 CC are useful as vectors for gene therapy which can be applied to cancer  
 CC therapy for destroying cancer cells selectively, the treatment of  
 CC autoimmune diseases and graft rejection reaction, and apoptosis induction  
 CC therapy for inflammatory cells in inflammatory diseases. Prior arts have  
 CC encountered the problem where if an adenovirus vector capable of  
 CC expressing an apoptosis-associated gene is introduced into animal cells,  
 CC the cells producing the virus will be destroyed because the period of  
 CC time required to induce cell death by apoptosis is shorter than that  
 CC required to replicate and produce the virus, resulting in failure to  
 CC obtain a recombinant virus having the integrated apoptosis-associated  
 CC gene. In this invention an apoptosis-resistant 293 cell line (having an  
 CC apoptosis resistant gene introduced) is established and overcomes the  
 CC problem. The present sequence represents the base sequence of the  
 CC plasmid pR-X1-bar, which contains the human Bcl-X1 gene, and  
 CC is used in an example from the present invention.  
 XX  
 XX Sequence 7372 BP; 2353 A; 1749 C; 1649 G; 1621 T; 0 other;  
 SQ  
 Query Match 2.8%; Score 106.8; DB 20; Length 7372;  
 Best Local Similarity 47.1%; Pred. No. 5.6e-14;  
 Matches 327; Conservative 0; Mismatches 367; Indels 0; Gaps 0;  
 QY 2 ttgggttcgctgttttcattcttctctccttaagggttaacaatgacagtaatt 61  
 DB 5133 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 5074  
 QY 62 catattgtgtaacagtgagttctgtgccaattatgtaacattcttgaattgttt 121  
 DB 5073 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 5014  
 QY 122 gtctcatgttttatttcttcttcttcttcttcttcttcttcttcttcttctt 181  
 DB 5013 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 4954  
 QY 182 tctactaagaagaattcagtttcttactagttcagtttcttcttcttcttcttct 241  
 DB 4953 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 4894  
 QY 242 ttttgggtgttttcttcttcttcttcttcttcttcttcttcttcttcttctt 301  
 DB 4893 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 4834  
 QY 302 ttaataacacataatatacaccattttgagcagcaataaatttgcatttcag 361  
 DB 4833 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 4774  
 QY 362 tacaattgtgagcaactcttcttcttcttcttcttcttcttcttcttcttct 421  
 DB 4773 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 4714  
 QY 422 ggtcaatacacaatgatacctaattgacccatttgaatttcgtttgaataatgct 481  
 DB 4713 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 4654  
 QY 482 agtacacacttattctgtatataatgaaagcgcaattctgtgaatttctgcat 541  
 DB 541 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT

DB 4653 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 4594  
 QY 542 ctgatttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 601  
 DB 4593 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 4534  
 QY 602 ctgcttgaagaacttgaatttgaattgattgattgattgattgattgattgatt 661  
 DB 4533 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 4474  
 QY 662 cttaagggaataacccaatgcacatacattc 695  
 DB 4473 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 4440  
 RESULT 8  
 ID AAX33180/c  
 ID AAX33180 standard; DNA; 7797 BP.  
 AC AAX33180;  
 XX  
 DT 25-JUN-1999 (first entry)  
 XX  
 DE Cowpox virus bsr full length gene sequence.  
 XX  
 KW Cowpox virus: bsr; viral vector; expression; apoptosis; resistance;  
 KW crm; bcl-2; bcl-x1, FLIP, survivin, IAP, ILP; adenovirus; cancer;  
 KW autoimmune disease; graft rejection reaction; inflammation;  
 KW inflammatory disease; ss.  
 XX  
 OS Cowpox virus.  
 XX  
 PN W09913073-A2.  
 XX  
 PD 18-MAR-1999.  
 XX  
 PF 07-SEP-1998; 98WO-JP04010.  
 XX  
 PR 08-SEP-1997; 97JP-0259235.  
 XX  
 PA (RRPG-) RPR GENCELL ASIA PACIFIC INC.  
 XX Hamada H;  
 XX WPI: 1999-243728/20.  
 DR New apoptosis-resistant virus-sensitive cell  
 XX  
 PS Example 1: Page 34-38; 51pp; English.  
 XX  
 CC The present invention describes an apoptosis-resistant virus-sensitive  
 CC cell line into which an apoptosis resistance gene has been introduced.  
 CC The recombinant viruses generated are capable of expressing apoptosis-  
 CC associated genes. These can then be used in a variety of diseases for  
 CC which the induction of apoptosis by gene transfer, or where the  
 CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses  
 CC are useful as vectors for gene therapy which can be applied to cancer  
 CC therapy for destroying cancer cells selectively, the treatment of  
 CC autoimmune diseases and graft rejection reaction, and apoptosis induction  
 CC therapy for inflammatory cells in inflammatory diseases. Prior arts have  
 CC encountered the problem where if an adenovirus vector capable of  
 CC expressing an apoptosis-associated gene is introduced into animal cells,  
 CC the cells producing the virus will be destroyed because the period of  
 CC time required to induce cell death by apoptosis is shorter than that  
 CC required to replicate and produce the virus, resulting in failure to  
 CC obtain a recombinant virus having the integrated apoptosis-associated  
 CC gene. In this invention an apoptosis-resistant 293 cell line (having an  
 CC apoptosis resistant gene introduced) is established and overcomes the  
 CC problem. The present sequence represents the cowpox virus bsr gene which  
 CC is used in an example from the present invention.  
 XX  
 XX Sequence 7797 BP; 2542 A; 1760 C; 1656 G; 1839 T; 0 other;











| Query Match           | 2.6%                                                          | Score 98.6        | DB 22    | Length 7442 |
|-----------------------|---------------------------------------------------------------|-------------------|----------|-------------|
| Best Local Similarity | 46.8%                                                         | Pred. No. 4.2e-12 |          |             |
| Matches 311           | Conservative 0                                                | Mismatches 354    | Indels 0 | Gaps 0      |
| QY 1                  | gtttggttcgcgtgttttcattcccttcctctcttaaggagggaataacaaatgaacgaat | 60                |          |             |
| Db 1909               | gtttattatttttagtttttttttaaaatttttttttttagatttttttttttttaatt   | 1968              |          |             |
| QY 61                 | tcatattgtgaacagcgcgcattccctgtgcacattatgacacattccttgtaattgt    | 120               |          |             |
| Db 1969               | tttttttttttttttagagtgctgtttttatttagtgttttttagatttttttagat     | 2028              |          |             |
| QY 121                | tgtttcagtttatttcatttccttccttccttttaggtgaacaaatccccaatca       | 180               |          |             |
| Db 2039               | tgtgttttttttttagatattttttttttttttttttttttttttttttttttttag     | 2088              |          |             |
| QY 181                | ttcacactagagaaatcacgttttatactagtttcagtttattattgtttattagt      | 240               |          |             |
| Db 2089               | ttatttttttttttagatttttttttttttttttttttttttttttttttttttaatt    | 2148              |          |             |
| QY 241                | gttttagttggtttccattatgtgatgcagaattatgaagggtgtgtgtgcgtgt       | 300               |          |             |
| Db 2149               | ttttatttttttttaaaatttttttttttttttttttttttttttttttttttttag     | 2208              |          |             |
| QY 301                | gttaataaacaataagattataacaccattttgcagtcacataaataatgaattca      | 360               |          |             |
| Db 2209               | tttaatttttttttagatttttttttttttttttttttttttttttttttttttag      | 2268              |          |             |
| QY 361                | gtacaatgtgcgaacatctcttcatttttttttttttttttttttttttttttttt      | 420               |          |             |
| Db 2269               | tattttttttttttaattttatttttttttttttttttttttttttttttttttttag    | 2328              |          |             |
| QY 421                | gggtgaataacaatgaactaaattatgcccatcttggaatttggtttgaaatatgac     | 480               |          |             |
| Db 2329               | tattttttttttttaaaattttatttttttttttttttttttttttttttttttttag    | 2388              |          |             |
| QY 481                | tagtaacaactatctctgtaattatgaaagacgaattctgtgtgaagtttgcat        | 540               |          |             |
| Db 2389               | tatttttttttttttagatttttttttttttttttttttttttttttttttttttag     | 2448              |          |             |
| QY 541                | tctgtatttttttttcaatttcttctctctgtgaaggttaacaactaagccaattat     | 600               |          |             |
| Db 2449               | tttttttttttttagtttttttttttttttttttttttttttttttttttttag        | 2508              |          |             |
| QY 601                | tctgtctagaacaacttagatttttgatgtgtttttttttttttttttttttttttt     | 660               |          |             |
| Db 2509               | tttttttttttttagattttatttttttttttaaaattctggttttttttttttag      | 2568              |          |             |
| QY 661                | tctttt 665                                                    |                   |          |             |
| Db 2569               | ttttt 2573                                                    |                   |          |             |

RESULT 15

ABL32527

ID ABL32527 standard; DNA; 7571 BP.

XX ABL32527;

XX

DT 26-MAR-2002 (first entry)

XX DE Human immune system associated gene SEQ ID NO: 500.  
 XX XX  
 KW Human; immune system associated; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antihaemic; cytosolic; noctropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antirheumatic; antidiabetic; antipsoriatic;  
 KW antileukemic; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 XX OS Homo sapiens.  
 XX PN WO200200928-A2.  
 XX PD 03-JAN-2002.  
 XX XX  
 XX PF 02-JUL-2001; 2001WO-EP07537.  
 XX PR 30-JUN-2000; 2000DE-1032529.  
 XX PR 01-SEP-2000; 2000DE-1043826.  
 XX PA (EPIG-) EPIGENOMICS AG.  
 XX PI Olek A, Piepenbrock C, Berlin K;  
 XX DR WPI; 2002-130909/17.  
 XX PT Nucleic acid comprising fragment of chemically modified gene, useful  
 XX PT for diagnosis and treatment of diseases associated with abnormal  
 XX PT cytosine methylation  
 XX PS Claim 1; SEQ ID NO 500; 32pp + Sequence Listing; German.  
 XX CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 XX SQ Sequence 7571 BP; 1640 A; 192 C; 2031 G; 3708 T; 0 other;

Query Match 2.6%; Score 98; DB 24; Length 7571;  
 Best Local Similarity 46.6%; Pred. No. 5.8e-12;  
 Matches 385; Conservative 0; Mismatches 435; Indels 6; Gaps 2;

QY 5 ggttgcgtgttttcaattccttcttcttaagggtlaatacgaatgaatcat 64  
 DB 2608 gattgcgtgttttcaattccttcttcttaagggtlaatacgaatgaatcat 64  
 QY 65 attgtaacagcgatctctgtaacattatgtaacattcttgcattgttgc 124  
 DB 2668 ttctcgaaagcttgcgttcttcttcttcttcttcttcttcttcttctt 2727  
 QY 125 tcatgtttatttcaatttc---ttactttttttaggttaaaacaaatgc 180  
 DB 2728 tctgtttttttgttttttttttttttttttttttttttttttttttttt 2727  
 QY 181 ttctactaaggaatcgaattcttactagttcagtttattatattatgaat 240  
 DB 2788 ttcttttgcgttttttttttttttttttttttttttttttttttttttt 2847  
 QY 241 gttttagttggttttctcatattgtaagaaatgaatgaatgaatgaatga 300  
 DB 2848 agtttttttcgttttttttttttttttttttttttttttttttttttttt 2907  
 QY 301 gtttaataacaaatgaattatataccattttgacgtcataaaatgaattca 360

DB 2908 ttttttagtttttttttttttttttttttttttttttttttttttttttttt 2967  
 QY 361 gtacaaattgycgaacactctctcaatttttttttttttttttttttttttttt 420  
 DB 2968 ttttttttttttttttttttttttttttttttttttttttttttttttttttt 3027  
 QY 421 ggttaataacaaatgaattatataccattttgacgtcataaaatgaatgaatga 480  
 DB 3028 ttatcgatatttttttttttttttttttttttttttttttttttttttttt 3085  
 QY 481 tagtaacacttctctgtatattataggaagaaagcgaattcttgcgaattgtcat 540  
 DB 3086 ttttttttttttttttttttttttttttttttttttttttttttttttttttt 3145  
 QY 541 tctgtatttttttttttttttttttttttttttttttttttttttttttttttt 3205  
 DB 3146 ttttttttttttttttttttttttttttttttttttttttttttttttttttt 3265  
 QY 601 tctgttttagaaactttgatttttttttttttttttttttttttttttttttttt 660  
 DB 3206 taccgttttttttttttttttttttttttttttttttttttttttttttttttt 3265  
 QY 661 tctttaagggaataacaaatgcgaactaatccatccattagaaactctttactca 720  
 DB 3266 tttttgttttttttttttttttttttttttttttttttttttttttttttttt 3325  
 QY 721 caaaaactcaacttttatatgcttatctgtgataattataaaagacagttctatca 780  
 DB 3326 ttttttttttttttttttttttttttttttttttttttttttttttttttttt 3385  
 QY 781 aattggtgcaactttatcatattttgttcaaatatttttttttttttttttt 826  
 DB 3386 atttttttttaagaattgttttttttttttttttttttttttttttttttttt 3431

Search completed: July 31, 2002, 14:11:05  
 Job time: 17362 sec

Thu Aug 1 08:29:56 2002

us-09-899-718a-1.rng

Page 14



Query Match 100.0%; Score 3785; DB 6; Length 3785;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3785; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtttggttcgctgttttcaatcttcttcttcaagggttaatacaatgaagaat 60  
DB 1 GTTGGTTGCGTGTGTTTCATTTCTTTCTTAAGGGGTAAACCAATGACAGTAAT 60  
QY 61 tcaatttgtaacagtgagatctcttgccaattatgacattctcttgtaattgtt 120  
DB 61 TCATATTGTGTACAGTGGATCTTGTCGAATTAATGACAAATTTCTTTGTAAATGTT 120  
QY 121 ttttcaagtattatctatcttcttcttacttttaaggtaaaacaatgcccgaattca 180  
DB 121 TGTTCATGTTTTATTTATTTCTTTACTTTTAGGGTAAACCAATGCCCAATTTCA 180  
QY 181 ttctaactaagagaatctcagtttatacagttcagtttattattattgttttaagt 240  
DB 181 TTTTACCTTAAGAGAAATTCAGTTTATTAACAGTTTCAAGTTTATTTATTTATTAAGT 240  
QY 241 gtttttagttggtttctcattatgltatgcaagaataaggggtgtgtgtgtgt 300  
DB 241 GTTTTATGTTGTTTCTCATTTATGTAATGATGAATATAGGGGTGTGTGTCGTGT 300  
QY 301 gtaataacaacaaatattatacaaccaatttttcagtcataaaatattgcaattca 360  
DB 301 GTTAATATATACACATATAGTATTTATACACCAATTTTGGACATCAATTAATATGCAATTTTCA 360  
QY 361 gtaacaatttgccgaacactctctcattcttcttcttcttcttcttcttcttctaa 420  
DB 361 GTACAAATTTGGCCAAACCTCTTCTTCATTTTATTTTATTTATTTTCTTTCTTTAA 420  
QY 421 gggtaataccaatgatactaatattatgcccatttggaaatttcgttttgaataatgac 480  
DB 421 GGGTAATACCAATGATCTTAATTTATGCTCATTTGGAAATTTCTGTTTGAATAATTAATGAC 480  
QY 481 tagtaacaactatcttctgatatataatgaaagcgaattctctgtgaattgtgtat 540  
DB 481 TAGTACACACTTATCTTGATATATATGAAAGGCAATTTCTGTGTAAAGTTTGTCTAT 540  
QY 541 tctgtatttttttcaattttcttctctctcttctgaaaggttaacaatgccaatattca 600  
DB 541 TCTGTATTTTTCATTTTCTTTCTTCTGGAAGGTAACACTTAATGCAATATTCAT 600  
QY 601 tctgtcttaagaacaacttagtattctgattgtgttgaatttattattcttcttct 660  
DB 601 TCTGTCTTAAGAAACCTTATGATTTTGAATTTGTTTATTTTATTTTCTTTTGTCT 660  
QY 661 tctttaaggaataacaaatgccaatcaatctcattcattagaanaactcttattctta 720  
DB 661 TCTTTAAGGAAATACCAATGCACTAATCCATCTTGAAGAAATCTCTTATCTTA 720  
QY 721 caaaaactcaacttllatalgcttalcgtgataatataaaagcaggttctattctta 780  
DB 721 CAAAAACCACTTTAATGCTTATTCGTGCAATTAATAAAGCACTATTTCTATCTTA 780  
QY 781 aattgctgcaaaacttatacatattgtcttaataatttttctcgaagatgattgacc 840  
DB 781 AATTGGCTGCAAACTTTATATCAATTTGTCTAAATTAATTTTCTTGAATGATGATAC 840  
QY 841 aatggcaactaatcattccgtgagcagcaatatacggaatgccaatgataattgtgt 900  
DB 841 AATGCCACTAATTCATTCGCGAGCAAGCAATATGSGGATGCGCATATTAATGATGCT 900  
QY 901 gtgcatttttcatctcgaagatgggaatgataccctacaaatgacacacacagcat 960  
DB 901 GTGCATTTTTCATCTCTCAAGCATGGGCAATGCAATACCTCAACATGACACACAGCAT 960  
QY 961 aacaaacacatgagcactcaagcagagacatgataacctgtctgcagacacaaagac 1020  
DB 961 ACACAACACATGAGCATGACGCGAGCATGCAATACCTGTGTGGCACACACACACAC 1020

QY 1021 cgacacacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1080  
DB 1021 CGACACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1080  
QY 1081 ggaactgtaactatactacacagcagcagcagcagcagcagcagcagcagcagcagcagc 1140  
DB 1081 GGACAGGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1140  
QY 1141 aggtgccaagaactagattttcaaggcagcagcagcagcagcagcagcagcagcagcagc 1200  
DB 1141 AGGTGCCAAGCACTAGTATTTTTCAGCGCAGCTGGGATTAATGCAAGGCTATTTTCTGT 1200  
QY 1201 tcttagagcgaagaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1260  
DB 1201 TCTTAGAGCAGAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1260  
QY 1261 caacacactcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1320  
DB 1261 CACACACTCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1320  
QY 1321 gatcgcagaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1380  
DB 1321 GATCGACAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1380  
QY 1381 acgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1440  
DB 1381 ACGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1440  
QY 1441 agcaaaagagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1500  
DB 1441 AGCAAAAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
QY 1501 aggaagcaatcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1560  
DB 1501 AGGAAGCAATCCCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1560  
QY 1561 agcgaagacaacatlgcccccaagaactgaagaagcagcagcagcagcagcagcagcagc 1620  
DB 1561 AGCGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1620  
QY 1621 ggcgtgacacacacacacacacacacacacacacacacacacacacacacacacacacac 1680  
DB 1621 GCGTGACATCACACACACACACACACACACACACACACACACACACACACACACACAC 1680  
QY 1681 agaagatgctgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1740  
DB 1681 AGAAGATGGGTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1740  
QY 1741 aaagcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1800  
DB 1741 AAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1800  
QY 1801 caagctgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1860  
DB 1801 CAGGCTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1860  
QY 1861 cgcctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1920  
DB 1861 CGCTGCGGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1920  
QY 1921 ctccactcaacgctcgctcttaagagagagcttcgctgcttaagaacactcaactctgtg 1980  
DB 1921 CTCACACTCAACGCTCGCTTTCAGGACGATGCTTGCGGCTTTTAACACCTTCTTGTG 1980  
QY 1981 tctatgacatgtgagcacaacagatlgctgtgccaatlgatccaaaggcaggtg 2040  
DB 1981 TCTATGACATGTAGCCCAACAGATGCTGCGCCACATGATGATCAAAAGCAGAGGTG 2040  
QY 2041 cctttaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2100  
DB 2041 CCTTTAAAGCAGCAGAGCTGCTGCCGCTTTCATTTACAGCGGCGCATGATGCGGGTGG 2100







```

Db 121 GGCATGAGACGTGCGAGCCGCGCCGCCAAAGCAGGAAACCGACCGATTTC 180
QY 3320 gaccggcggctgctctcaatggtagtgcggcgaaggcgagcgatgaactctgt 3379
Db 181 GACCGCGGCTGCTCTCCATTGGTGGTGGCGCCACGCGGCGGCGGATGAACCTCGTG 240
QY 3380 ttctgcgcgcgcagatgagccctcgggaagaactgagcgctcggcgagcgtctcggg 3439
Db 241 TTCTGTCGGCGCGAGATGGCGCCCTGGAGCAAGACTGGCGGCTCGGCGAGCTCTCGGG 300
QY 3440 ggcctcccgccgcagatgagcgtaagcttgcgccactgctctataaagtctctcc 3499
Db 301 GGCCTCCCGCGCGCATGGCCCTTAAGTGGCGCCACTGCTCTTAATAAGTTCTTCC 360
QY 3500 tgcagcaatgctgcctgcttaacaagcggtgcgtctcgtcgaagccaaagtcacgggt 3559
Db 361 TGCAGCCATGCTCGCTTACCAAGCGGTGCGCTGCTCGAGGCAACGGGTCAACCGGGT 420
QY 3560 catgtcatctcccgcgctacgaacgaatataagaagccctcgggaacacagcgtcatctc 3619
Db 421 CATGTCATCTCCCGCGCTAGACACTACAGACGCCCTGGAGACACCGTCATCTC 480
QY 3620 cgaagtatatatcgccacatgaattatcaaatcacatgctctcgtacatatctgcaa 3679
Db 481 CGAGGTATATATCCGCCACATGAATTAACAAATTCACATGCTCTCTGCAATTTCTGCAA 540
QY 3680 gacttactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 3739
Db 541 GACTTACTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
QY 3740 gtacttactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 3785
Db 601 GTACTTCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 646

RESULT 4
AB029063 2781 bp DNA linear PLN 21-DEC-1999
LOCUS Triticum durum waxy gene for starch synthase (GBSSI), complete cds.
DEFINITION AB029063
VERSION A029063.1 GI:6624284
KEYWORDS starch synthase (GBSSI).
SOURCE Triticum durum DNA.
ORGANISM Triticum durum subsp. durum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
REFERENCE 1 (sites)
AUTHORS Mural, J., Taira, T. and Ohta, D.
TITLE Isolation and characterization of the four Waxy genes encoding the
granule-bound starch synthase in tetraploid wheats
JOURNAL Appl. Biol. Sci. (1999) In press
AUTHORS Mural, J., Taira, T. and Ohta, D.
TITLE Direct Submission
JOURNAL Submitted (16-JUN-1999) Jun Murali, Osaka Prefecture University,
College of Agriculture; 1-1 Sakuen-cho, Sakai, Osaka 599-8531,
Japan (E-mail:junkiedmeter.plant.osaka-fu-u.ac.jp,
Tel:81-722-54-9409, Fax:81-722-54-9409)
FEATURES
source
Location/Qualifiers
1..2781
/organism="Triticum durum subsp. durum"
/db_xref="taxon:4567"
join(1..321,404..484,569..667,777..930,1056..1156,
1256..1609,1701..1880,1976..2167,2258..2344,2443..2571,
2655..2781)
/gene="waxy"
join(1..321,404..484,569..667,777..930,1056..1156,
1256..1609,1701..1880,1976..2167,2258..2344,2443..2571,
2655..2781)
/gene="waxy"
/codon_start=1

```

```

/product="starch synthase (GBSSI)"
/protein_id="BA8511.1"
/db_xref="GI:6624285"
/translation="MALVTSQLATSGVLTVDREPRPFGGLRPNDALGMRT
VGAAPKQSRKPRGRNRCGLSMVATGATVATVGAEMAPSKKGGIDVGL
PAMAMNGHVMVLSRVDYDADMTSVISLKYVDREPRFTRCKRQVDYFD
HRCFLEKVRGKTEKLYGPDAGTIDEDNQRSLCOALVYRILDLNNPHEGYP
GEDVAVCDMHTGLACYLKSNSNGYRPAKVPCTIHSYSGRESFDPFALNL
PDKFSFDFIDGYDPVEGRKINMKAGILQADKVLVSPYAEELISGEARCELD
NMRLTGILVNGMDVSEMDPIKDFELVDVTYTALEKALKEALQAEGLPVDK
KVLVAFIGRLEOKGPDVMAIPEIYVEEVOYVILGTGKKREPLKSVBEKFT
KYRAVYRENAPLAHOMAGADYLAITSREPGCLQLOGMRGTFCACASTGGLDVI
VSKTGFHNGRLSVDCNVYEPADVKVVTTLERAVKVVGTAPVHBMVNCMIODLSMK
GPAKMEVDLLEIGVSGSPGIYGEIALALENVAAP"

BASE COUNT      569 a      810 c      820 g      582 t

Query Match      16.9%; Score 638; DB 8; Length 2781;
Best Local Similarity 99.2%; Pred. No. 1,6e-111;
Matches 641; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3140 atgagcgtctgctacgctccagctcgcacactcgcgacccgtctcagcgctaccgac 3199
Db 1 ATGGCGGCTGTGTTCACTCCAGCTCCAGCTCCGACCTCGGACCGTCTCAGCGTCACCGAC 60
QY 3200 agattccgctcagagtttccagggccttgaagcccccggaacccggatgacggcctc 3259
Db 61 AGATTCCGGCGCTCAGAGTTTTCAGGGCTGAGGCCCCGGAACCCGGGATGCGGCGCTC 120
QY 3260 ggcataagagctgctgagagcgccgcgcccaagaagaagaacacgacgattc 3319
Db 121 GGCATGAGACTGTGCGAGCGACCGCGCCGCCAAAGCAAGAGAAACCGACCGTGGG 180
QY 3320 gaccgagctgctccctccatggtggtggtgcgcaagcgagcgagcgatgaactctgt 3379
Db 181 AACCGCGGTGCTCTCCATGATGTTGCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCTG 240
QY 3380 ttctgcgcgcgcagatgagccctcgggaagaactgagcgctcggcgagcgtctcggg 3439
Db 241 TTCTGTCGGCGCGAGATGGCGCCCTGGAGCAAGACTGGCGGCTCGGCGAGCTCTCGGG 300
QY 3440 ggcctcccgccgcagatgagcgtaagcttgcgccactgctctataaagtctctc 3499
Db 301 GGCCTCCCGCGCGCATGGCCCTTAAGTGGCGCCACTGCTCTTAATAAGTTCTTCC 360
QY 3500 tgcagcaatgctgcctgcttaacaagcggtgcgtctcgtcgaagccaaagtcacgggt 3559
Db 361 TGCAGCCATGCTCGCTTACCAAGCGGTGCGCTGCTCGAGGCAACGGGTCAACCGGGT 420
QY 3560 catgtcatctcccgcgctacgaacgaatataagaagccctcgggaacacagcgtcatctc 3619
Db 421 CATGTCATCTCCCGCGCTAGACACTACAGACGCCCTGGAGACACCGTCATCTC 480
QY 3620 cgaagtatatatcgccacatgaattatcaaatcacatgctctcgtacatatctgcaa 3679
Db 481 CGAGGTATATATCCGCCACATGAATTAACAAATTCACATGCTCTCTGCAATTTCTGCAA 540
QY 3680 gacttactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 3739
Db 541 GACTTACTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
QY 3740 gtacttactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 3785
Db 601 GTACTTCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 646

RESULT 5
AF474373 124050 bp DNA linear PLN 12-FEB-2002
LOCUS Hordeum vulgare BAC 259116, complete sequence.
DEFINITION AF474373
VERSION AF474373.1 GI:18652401

```



```

75159..75245,75343..75471,75548..75664)
/gene="259116.5"
/codon_start=1
/product="granule-bound starch synthase"
/protein_id="AAI77109.1"
/db_xref="GI:18652407"
/translation="MAALATSQLATSGVLGVTRFRPRPGFGRPRNPADALGMPT
IGASAPKROSKRAHGRSLSVVRATGSGMNVFVGAEMAPSKTGGLDVLGCLP
PAMANGHRVAVSPRYDOXDAMDTSVTSIKADIERERFHCYKRGVDFYIDH
EDVFCNDMHTGLLACTKSNYSNGIYRPAKVAFCIHNSYGRFSFDFALNLP
DRKSSDFIDGDKPVEGRKINMMKAGIIOADKVLVSPYAEELISGEARGCELN
IMRLTGITGVNGDVSEMDPKDFLAVNDITLAEAKLNEALQAEGLVDK
VPLVAFIRLEEEKGPDVMAIPEIKEDVOIILGTGKKREKILKSMKEFPK
VRAYVENAPLAHOMAGADLAVTSRPEPVOIIOLOMRGTFCVASCGLVDITV
EGKGFHMGRLSDCNVPEADYKVVITTKRAVKVGTPTQEMVKNMODLSWKG
PAKNMVDLLELGESESPGIVGEELIAPLAEVNAAP"
74093..74118
/rpe_unit-at
complement(join(<77035..77091,78213..78398,78548..78852,
7967..80040,80132..80247))
/gene="259116.6"
/product="unknown"
/complement(join(<77035..>80247)
/gene="259116.6"
7967..80040,80132..80247))
/gene="259116.6"
/codon_start=1
/product="unknown"
/protein_id="AAI77110.1"
/db_xref="GI:18652408"
/translation="MASSARAAASARSALRRAPLAASSSSSAAGAAILRASAAG
LATVPLHSAARLGSCTRDGNNAFSNHRMSNKKHPTSIDPEIETAPFDLIAA
SMNELPSIIVEEAKKEVSKATDDKAGEALENPRAEAEESGLVTLRMALDDIC
GLTEGNSGPIPGYLEAVKSAISHYMTLTSFGEHYLKKVETELGKIMHLKMC
SGTSEWKGITLITSGISYEMRA"
80943..82286
/feature="repetitive sequence RP-2"
82378..82598
/transposon="MITE"
83418..84091
/feature="repetitive sequence RP-2"
85127..85346
/transposon="MITE"
join(<91421..91518,91736..>92153)
/gene="259116.7"
/product="hypothetical protein"
<91421..>92153
/gene="259116.7"

```

Query Match 14.8%; Score 560.2; DB 8; Length 124050;  
Best Local Similarity 65.9%; Pred. No. 2.8e-96;  
Matches 1106; Conservative 0; Mismatches 418; Indels 155; Gaps 14;

```

Db 72202 CCGCCCGCCGTCCTCCGTCACAGAGAGAGMAGAACAGTTTCAGCAAGAAGAGATC 72261
QY 2430 agacaggtaagcagagacagataagtaagcggcgccagttccggcgccggagcag 2489
Db 72282 AGATC-----GATCAGAGTACGATTTCCCATGCCGCCGCGGATCCGCGAGTACA 72317
QY 2490 ggalagatcgattagtcggtcctcaaatcaagtcggtgctcagtagatagatagac 2549
Db 72318 TCAGATGGTTTATTAGTTGGCCACCACTCGTTGGTTGCATTCTAGTACATCCAG 72377
QY 2550 catcaaatccggcagatggttagatccagagtccttccttctaactaaagtcgag 2609
Db 72378 CATTCCTGTGGCGCCCGCTGTGAGATCCACTGTCCTTTTGTTCGACTTCGTCGCG 72437
QY 2610 agcgttaagt---tgaagatcttcctatagatcgatagatcaatcaatgataaataa 2666
Db 72438 CAACTGACTATCCATGGATCTTCTTATACATTCCTGGATCCAAATCTCGATGTA----- 72492
QY 2667 aaaaaaagatttaaaatcattctactgctagctagatgagatgattctatgtgaacgat 2726
Db 72493 -----CTATGATGATTCCTCTCGCAACAGATCTTA 72522
QY 2727 gatctcgagacagatcca---atgattcatgcccgcctagaggttaataacagtag 2782
Db 72523 GATTTCAGAAACAGATCCACGATACCGCTTCATGCAATGTTCCGGAATCATTTAAGGCTTG 72582
QY 2783 acagagcagacataatgacgcaataaacattctgttcttagccgagttgagcaaca 2842
Db 72583 ACACAGGGAACATACAT-----AAGAAATTCGTGCTGTGATGTTTCTTTCTAAGAA 72636
QY 2843 gtcagtcgacgacccaagccttgattttgtttgtttgttggtggtggtcagtcgc 2902
Db 72637 AGGAGCTAATATATATGTTTCGTCATTT----- 72665
QY 2903 accctacagacaattcattctctcagccagtcaccgccgtgcagcgatttaagc 2962
Db 72666 -----AATATATTAA 72676
QY 2963 ttaattactacagtcgagagagttcatatatactctgltacgttaatttga 3022
Db 72677 AGAAGAAATTTCCCTTTAATGACGAAACCGGACAAACCGGTACTTATGTTGG 72736
QY 3023 ttcaattcaaatgttaaatccagaagaactgacgcaaatctcg-----ttaactc 3077
Db 72737 AATTCGGAAGAAATGTAAGTGCACAAACCTGACTGAAATTCGATATTATTCCTTC 72796
QY 3078 actact--cactaacaatagtcagtcgtctctctgctcagtagtagcacaccctgcgc- 3134
Db 72797 ACTACTCACACTGACGATCAGTGCAGTCTCTCTACTCAGTAGCCACACCTGTGTGCG 72856
QY 3135 --ggccatgagcggtcctctgtaagtcctcagctcgcaactcgcaacctctcaagct 3192
Db 72857 CGCGCGCATGGCGGCTCTGCGCACGTCGCCACGTCGCCACCTCGCACCGCTCGGCGT 72916
QY 3193 caacgacagatccggtcggtccaggttttcaaggacctgaagcccggaagcccgagatgc 3252
Db 72917 CACCGACAAATTCGCGGCTCCAGGTTTTCAGGCGCTCAGGCCCCGGAACCGCGGAGTGC 72976
QY 3253 ggcgtctggcatgagactgttcgagcgagcgcccccaaaacaaacagaaacacgca 3312
Db 72977 GGGGCTTGTATGAGACATATCGAGCAAGCGCGCCCGCGAAGACCGGAAAGCGCA 73036
QY 3313 cggattcgacggggtgtgctctcatggtgtgtgcgcgaagggcagcgagcatgaa 3372
Db 73037 CCGCGGAGCCGCGGCTCTCTCGTGTGTGTGCGCCCAAGGCGCA--CGGGATCAA 73093
QY 3373 cctcgtgtcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3432
Db 73094 CCGCGGTTTCGTCGCGCGCGGAGATGCGCCCTCGAGCAAGACGCGCGCTTGGCGAGT 73153
QY 3433 cctcgggggctctcccgccgcacatgagcggttaagcttgccgcaactgc-----cttctat 3486
Db 73154 CCTCGCGGCGCTCCCGCGCGCATGCGCGTAAAGTACATCAGACGACCAACCAACCAT 73213

```















QY 3641 gaattcacaattcactgctcctgacattctgcaagacttactgactgctgact 3700  
 Db 538 GAATTCACAGAGTTCACATGCTCCTGCACATTTTCAGAGGTTCCACTACCGAGTGAAT 597  
 QY 3701 ctgcgcatcaagtgctgctgcaagtgtagagggatgagttactctccactgtctacaagcg 3760  
 Db 598 TTCACAGATCAAGGTGCGGACGAGTACGAGGAGGAGTACTTCCACTACAGG 657  
 QY 3761 cggggtgacccgctgctgctgac 3785  
 Db 658 CGGGGTGACCGCGTCTGCTGCAC 682  
 RESULT 12  
 AF110374 2826 bp DNA linear PLN 12-NOV-1999  
 LOCUS Aegilops speltoides granule-bound starch synthase wx-Tsb protein  
 DEFINITION (wx-Tsb) gene, complete cds.  
 ACCESSION AF110374  
 VERSION AF110374.1 GI:6318539  
 KEYWORDS Aegilops speltoides.  
 SOURCE Aegilops speltoides.  
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Aegilops.  
 REFERENCE 1 (bases 1 to 2826)  
 AUTHORS Yan, L., Bhawe, M., Fairclough, R., Konic, C., Rahman, S. and Appels, R.  
 TITLE The genes encoding granule-bound starch synthases at the waxy loci  
 of the A, B and D progenitors of common wheat  
 JOURNAL Genome (1999) In press  
 REFERENCE 2 (bases 1 to 2826)  
 AUTHORS Yan, L., Bhawe, M., Fairclough, R., Konic, C., Rahman, S. and Appels, R.  
 TITLE Direct submission  
 JOURNAL Submitted (02-DEC-1998) School of Life Sciences and Technology,  
 Victoria University of Technology, Werribee Campus, PO Box 14428  
 MCMC, Melbourne Victoria 8001, Australia  
 FEATURES  
 source  
 1. 2826  
 /organism="Aegilops speltoides"  
 /cultivar="AUS 21638"  
 /db\_xref="taxon:4573"  
 join(<32, 355,455, .535,625, .723,837, .990,1116, .1216,  
 1316, .1669,1754, .1933,2016, .2207,2291, .2377,2466, .2594,  
 2710, .>2826)  
 /gene="wx-Tsb"  
 <32, .>2826  
 /gene="wx-Tsb"  
 /note="waxy"  
 join(32, 355,455, .535,625, .723,837, .990,1116, .1216,  
 1316, .1669,1754, .1933,2016, .2207,2291, .2377,2466, .2594,  
 2710, .2826)  
 /gene="wx-Tsb"  
 /note="GBSSI"  
 /product="granule-bound starch synthase wx-Tsb protein"  
 /protein\_id="AF06937.1"  
 /db\_xref="GI:6318540"  
 /translation="MAALVTSQLATSATVLTGIDRRHAFQGVRRSPADAPLIGRT  
 VGASAAFKQSRRAHGRTRCLSMVRAVSGSMNVFVGAEMAPSKGGIGDFG  
 LPPAMANGHRVWVTSIPRDYDKAMDTSVSGIKADVERERYHCKRKRDYRV  
 DHCPLFEKVGKTKETKIGDADCTEDENDLRSLCQALEPRLIDNNPYSFG  
 YGDDVYFVCDMRTGLLACTLKSNTQSNIGITAVKFCIHNTSYGREGSPDPAOLN  
 LPRFSPEDFDIGDKPVGKRINMKAGIIOADVLTSPVYAEELISGARGCEL  
 DNTMLRIGITGVANGADSEMDPTKDNFLAANNITTALEGLANKLEALQAEVGLPD  
 RKVLAIFGRLEBQKGPVMIASIEPIKEEDVQIVLGTGKKRERLKGVEKFP  
 SKYRAVRENAPLAHOMAGADVLAVTSPEEDVQIVLGTGKKRERLKGVEKFP  
 IVEGKTGFHGRSLSDVCNVEPADVKKVYTTLLKRAKVVCTAYHEMVA NCMIQDLSW  
 KGPKNMEDVLELTVGSESPVTEELIPLAMENVAAP  
 BASE COUNT 581 a 845 c 815 g 585 t

Query Match 12.68; Score 478.2; DB 8; Length 2826;  
 Best Local Similarity 87.28; Pred. No. 5.3e-81;  
 Matches 613; Conservative 0; Mismatches 63; Indels 27; Gaps 7;  
 QY 3109 ttgtgcagagtagcacaacccctgagcgcacatggcggtctgttcaagtcacagctcgc 3168  
 Db 1 TTGCTGCAGAGTAGCCACACCTTGCTGCGGACCATGCGGCTGTGTACGTCCAGCTCGC 60  
 QY 3169 caactcggacacgcttcctcagcgtcaccagaagaatctcggcgttcaagtttcaagacct 3228  
 Db 61 CACCTCCGACACCGTCTCTCGGCATCACGACAGGTTCGGCATGACAGTTTCCAGGGCGT 120  
 QY 3229 gaggcggcgaacccggcgatgctgcgcctcgcgcataagagatgtagagcagcgcgc 3288  
 Db 121 GAGGCCCGGGAGCCCGGACAGTGGCGGCTGGCATGAGCATGCTCGACGACGCCCGC 180  
 QY 3289 cccaaag---caagcagagaacccgacacgattcgaccggcggtgctcctcctcagtggt 3345  
 Db 181 CCCGACGACACAAAGCCGGAAGGACACCCGGGACCCGGCGGTGCTCTGATGGTGT 240  
 QY 3346 ggcgcgaaggcgagcgcgcatgaacctgctgttctgcgcgcagagatgagcgcctg 3405  
 Db 241 GCGGCGCACCGGCGGAGCGGCGCATGAACTCGTGTGTCGGCGCGAGATGGCGCCG 300  
 QY 3406 gagcaagacgagcgctcgcgcagcgtccctcggggggtcctcccgccatgagcgta-- 3463  
 Db 301 GAGCAGACCGGGGCGCTCGGCAAGCTCTCGGGGGGCTCCGCCAGCATGAGCCCTTAAG 360  
 QY 3464 -----agcttgccacactccttctataa--atgttctcctcagcagcagctgc 3515  
 Db 361 CTAGCTAGCTAGACACACCTCTCTCTCATATATTTCTTCTTGACGACACCATGCTTGGC 420  
 QY 3516 -----gtacaaggggtgcgttcgtcgtgacgcaacggtacacggtatccaggtat 3568  
 Db 421 ATTACAGTAGTACACATGATGCTGTGTCGACGAGCAACGAGTACCGGATGATCAT 480  
 QY 3569 ctcggcgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3625  
 Db 481 CTCGCCGCTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 540  
 QY 3626 -ataatccgcacatgaattat--cacaattacatgtctcctcagcattctcgaagc 3682  
 Db 541 CATATATCCGCCACATGAAATCATCCAAAATTCATATGCTCTCGACATTTCTGCAAGT 600  
 QY 3683 ttactgactggtgtagctcgcagatcaagtcggttgcaggtacagaggtgtagtga 3742  
 Db 601 TCTACTACCGACTGAT--TTGCAGATCAAGGTGCGGCTGATGACGAGAGGTGAGGTA 659  
 QY 3743 ctccactgtacaaagcgcggtgtagcgcgtgttcgtcgcag 3785  
 Db 660 CTTCATTTGTATCAAGCGCGGGGTGACCGCGGTGTTGCTGAC 702  
 RESULT 13  
 AB029062 2793 bp DNA linear PLN 27-JAN-2001  
 LOCUS Trilicium turgidum subsp. dicoccoides waxy gene for starch synthase  
 DEFINITION (GBSSI) complete cds.  
 ACCESSION AB029062  
 VERSION AB029062.1 GI:6624282  
 KEYWORDS starch synthase (GBSSI).  
 SOURCE Trilicium turgidum subsp. dicoccoides  
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Poideae; Triticeae; Triticum.  
 REFERENCE 1 (sites)  
 AUTHORS Mural, J., Talra, T. and Ohta, D.  
 TITLE Isolation and characterization of the four waxy genes encoding the  
 granule-bound starch synthase in tetraploid wheats  
 JOURNAL Appl. Biol. Sci. (1999) In press  
 REFERENCE 2 (bases 1 to 2793)





Thu Aug 1 08:29:55 2002

us-09-899-718a-1.rge

---